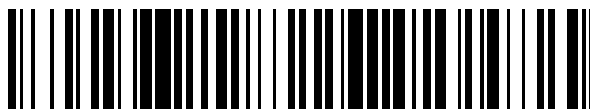


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 735 015**

51 Int. Cl.:

C12Q 1/6869 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **26.11.2014** **PCT/US2014/067582**

87 Fecha y número de publicación internacional: **04.06.2015** **WO15081178**

96 Fecha de presentación y número de la solicitud europea: **26.11.2014** **E 14830903 (2)**

97 Fecha y número de publicación de la concesión europea: **01.05.2019** **EP 3074534**

54 Título: **Composiciones y métodos para secuenciar polinucleótidos**

30 Prioridad:

26.11.2013 US 201361909316 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

13.12.2019

73 Titular/es:

**ILLUMINA, INC. (100.0%)
5200 Illumina Way
San Diego, CA 92122, US**

72 Inventor/es:

**STAVA, ERIC;
GUNDLACH, JENS, H.;
MANDELL, JEFFREY, G.;
GUNDERSON, KEVIN, L.;
DERRINGTON, IAN, M. y
MOHIMANI, HOSEIN**

74 Agente/Representante:

ELZABURU, S.L.P

ES 2 735 015 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Composiciones y métodos para secuenciar polinucleótidos

Campo de la invención

La presente descripción se refiere en general a métodos y composiciones para caracterizar un polinucleótido diana, incluyendo la caracterización de la secuencia del polinucleótido diana.

Como la información codificada en cada polinucleótido (por ejemplo, ADN o ARN) es de suma importancia para la medicina y las ciencias de la vida, existe la necesidad de secuenciar los polinucleótidos de forma rápida y económica. En la actualidad, las técnicas de secuenciación comerciales requieren de la preparación de las muestras y de las bibliotecas, siendo laboriosas ambas preparaciones. Además, las lecturas son más lentas de lo deseado para muchas aplicaciones. Por lo tanto, el rendimiento es limitado y los costes son relativamente altos. La secuenciación de nanoporos representa un nuevo método que se está desarrollando para secuenciar de forma rápida y económica cualquier polinucleótido diana.

La secuenciación de nanoporos utiliza un nanoporo, que puede proporcionar un canal para una corriente eléctrica iónica. El polinucleótido es impulsado electroforéticamente a través del nanoporo y, cuando el polinucleótido pasa a través del nanoporo, reduce la corriente eléctrica a través del nanoporo. Cada nucleótido que pasa, o una serie de nucleótidos, produce una corriente eléctrica característica, y el registro de los niveles de corriente corresponde a la secuencia del polinucleótido. Dado que algunos niveles de corriente se rigen por múltiples nucleótidos (generalmente 3-4), sigue existiendo la necesidad de mejorar el estado de la técnica para mejorar las precisiones. Cualquier información adicional sobre los niveles de corriente obtenidos, a medida que el polinucleótido se transloca a través del nanoporo, como la forma y la duración, puede proporcionar ventajas.

Un desafío común para la secuenciación de nanoporos es que la translocación del polinucleótido a través del nanoporo es tan rápida que los niveles de corriente para nucleótidos individuales son demasiado cortos para ser resueltos. Un enfoque para la secuenciación de nanoporos implica la translocación controlada de un polinucleótido a través del nanoporo bajo la guía de una proteína de unión a polinucleótido, como una helicasa, translocasa o polimerasa, contra un potencial de voltaje. A pesar de esta translocación controlada, todavía existen varios modos de error en la secuenciación que contribuyen a una pobre precisión en la secuencia. El documento WO2013/057495 describe un método enzimático relacionado con un método para caracterizar un polinucleótido diana.

Por lo tanto, existe la necesidad de conseguir métodos y composiciones que proporcionen una translocación controlada adicional del polinucleótido a través del nanoporo y una mejor resolución de la translocación de nucleótidos en la discriminación de nucleótidos. La presente descripción satisface esta necesidad y proporciona ventajas relacionadas.

Sumario de las realizaciones

Se proporciona un método para caracterizar un polinucleótido diana de acuerdo con las reivindicaciones.

La caracterización del polinucleótido diana incluye la identificación de uno o más de: (1) la secuencia del polinucleótido diana; (2) la modificación del polinucleótido diana; (3) la longitud del polinucleótido diana; (4) la identidad del polinucleótido diana; (5) la fuente del polinucleótido diana, o (6) la estructura secundaria del polinucleótido diana. También se proporciona un método para modular un paso de translocación fraccional de un polinucleótido diana a través de un poro y una composición para caracterizar un polinucleótido diana, que comprende un poro, una helicasa Hel308 y un polinucleótido diana contenido en una solución de menos de 1 mM de ATP o solución de un análogo de nucleótido.

Breve descripción de los dibujos

La Fig. 1A muestra el modelo electrostático de un gusano medidor para la translocación de un polinucleótido por una helicasa.

La Fig. 1B ilustra esquemáticamente una primera composición ejemplar que incluye un poro en contacto con una helicasa Hel308, de acuerdo con algunas realizaciones.

La Fig. 1C ilustra esquemáticamente las etapas en un método para caracterizar un polinucleótido diana.

La Fig. 2A muestra una comparación de los eventos de translocación de helicasa Phi29 polimerasa y Hel308 Tga, de acuerdo con algunas realizaciones. Las etapas de translocación fraccionarias observadas con una helicasa Hel308 Tga se muestran en comparación con las etapas de translocación observadas con una ADN polimerasa phi29.

La Fig. 2B muestra una comparación de los eventos de translocación de polimerasa Phi29 y helicasa Hel308 Tga, de acuerdo con algunas realizaciones. Las etapas de translocación fraccional observados con una helicasa Hel308 Tga se muestran en comparación con los niveles de corriente pronosticados, generados por una plantilla de polinucleótidos de hebra única que se translocan a través de un nanoporo MspA-M2 utilizando una polimerasa Phi29 como el motor

molecular, con los observados utilizando una helicasa Hel308 Tga como motor molecular.

La Fig. 2C muestra una comparación de los eventos de translocación de la polimerasa Phi29 y la helicasa Hel308 Tga, de acuerdo con algunas realizaciones. Las etapas de translocación fraccionarias observadas con una helicasa Hel308 Tga se muestran en comparación con las etapas de translocación observadas con una ADN polimerasa phi29 para una secuencia de nucleótidos repetida simple (SEC ID NO: 74).

La Fig. 3 muestra un mecanismo propuesto "basado en sujeción" para un paso de translocación fraccional, de acuerdo con algunas realizaciones.

Las figuras 4A y 4B muestran el efecto ejemplar de la concentración de ATP en el tiempo de permanencia de las etapas de translocación fraccional, de acuerdo con algunas realizaciones.

La Fig. 5 representa la precisión de la reconstrucción de la secuenciación (Modelo de Markov oculto (HMM)) para trazas de corriente generadas in silico de etapas completas (diamantes) y medias etapas (cuadrados) (descritas a continuación) con varios niveles de ruido agregado, de acuerdo con algunas realizaciones.

La Fig. 6A representa transiciones de estado con probabilidad no nula necesaria para que un HMM decodifique la secuencia en un nanoporo en el que el polinucleótido es movido por una enzima motora, de acuerdo con algunas realizaciones. El motor es phi29 DNAP o un polinucleótido de traslación de enzima similar en etapas para 1 nucleótido.

La Fig. 6B representa las transiciones de estado con una probabilidad no nula necesaria para que un HMM decodifique la secuencia en un nanoporo en el que el polinucleótido es trasladado por una enzima motora, de acuerdo con algunas realizaciones. El motor es la helicasa Hel308 o una enzima similar que permite el movimiento fraccional del polímero.

La Fig. 7 representa la precisión esperada de encontrar patrones actuales en función del desplazamiento gaussiano, de acuerdo con algunas realizaciones. Los diamantes representan un motor que tiene un escalón de nucleótido completo. Los círculos representan un motor que tiene una etapa de translocación fraccional, y los cuadrados representan un motor que tiene una etapa de translocación fraccional combinada con valores de duración.

La Fig. 8 muestra una modulación ejemplar de la actividad de helicasa Hel308 con concentraciones variables de pirofosfato, de acuerdo con algunas realizaciones.

La Fig. 9 muestra la modulación ejemplar de la actividad de helicasa Hel308 con el inhibidor de nucleótidos ortovanadato de sodio y con el análogo de nucleótido hidrato de sal de adenosina 5'-(β,γ -imido)-trifosfato de litio, de acuerdo con algunas realizaciones.

La Fig. 10 muestra un ejemplo de un método para usar la información proporcionada por la etapa de translocación fraccional adicional que se puede obtener a partir de dos lecturas de secuencia independientes, usando niveles y duración de nivel, de acuerdo con algunas realizaciones.

La Fig. 11 muestra un ejemplo de un método para usar la información proporcionada por la etapa de translocación fraccional adicional que se puede obtener a partir de dos lecturas de secuencias concurrentes, usando niveles y duración de nivel, de acuerdo con algunas realizaciones.

La Fig. 12 muestra un ejemplo de un método para usar la información proporcionada por el etapa de translocación fraccional adicional utilizando trazas de corriente, con o sin información sobre la duración, de acuerdo con algunas realizaciones.

Las figuras 13A-13E muestran la translocación de polinucleótidos controlada por helicasa Hel308 basada en un complejo de polinucleótidos ternario con un sitio de unión saliente 3' de helicasa Hel308 y un ancla bicapa de colesterol, de acuerdo con algunas realizaciones. Los círculos rellenos (●) indican un fosfato de 5'. El diamante relleno (◆) indica un colesterol 3'. Un círculo semitransparente lleno de marcas indica la helicasa Hel308. Las líneas de puntos indican una longitud arbitraria. Las flechas grises grandes indican la dirección del movimiento del polinucleótido (con o contra el campo aplicado) del polinucleótido dentro o fuera del poro. Las flechas negras grandes indican la dirección de la translocación de la helicasa a lo largo del polinucleótido, que va de 3' a 5'. El poro (objeto cónico en forma de embudo) se asienta en la membrana (líneas horizontales dobles).

Las figuras 14A-14D muestran la translocación de polinucleótidos controlada por Hel308 helicasa basada en un complejo de polinucleótidos ternario con un sitio de unión saliente de 3' de helicasa Hel308 y un anclaje bicapa de colesterol, de acuerdo con algunas realizaciones. Círculo relleno (●) indica un fosfato de 5'. El diamante relleno (◆) indica un colesterol 3'. Un círculo semitransparente lleno de marcas indica la helicasa Hel308. Las líneas de puntos indican longitud arbitraria. Las flechas grises grandes indican la dirección del movimiento del polinucleótido (con o contra el campo aplicado) del polinucleótido dentro o fuera del poro. Las flechas negras grandes indican la dirección de la translocación de la helicasa a lo largo del polinucleótido, que va de 3' a 5'. El poro (objeto cónico en forma de embudo) se asienta en la membrana (líneas horizontales dobles). Los símbolos son los mismos que en las Figs. 13A-13E. En este esquema, hay un único polinucleótido de hibridación "iii" que crea un saliente de 3' en el polinucleótido "ii" para que la helicasa Hel308 se adhiera, y también contiene un resto de colesterol opcional.

- Las figuras 15A-15C muestran la translocación controlada en la misma dirección que la fuerza de gradiente, de acuerdo con algunas realizaciones. Un círculo semitransparente lleno de marcas indica la helicasa Hel308. Las líneas de puntos indican longitud arbitraria. Las flechas grises grandes indican la dirección del movimiento del polinucleótido con el campo aplicado en el poro. Las flechas negras grandes indican la dirección de la translocación de la helicasa a lo largo del polinucleótido, que va de 3' a 5'. El poro (objeto cónico en forma de embudo) se asienta en la membrana (líneas horizontales dobles).
- La Fig. 16 ilustra esquemáticamente varios motivos (SEC ID NOS 75-81, respectivamente, en orden de aparición) que se han identificado en la familia SF2, por ejemplo, las helicasas DEAD-box (SEC ID NO: 2), de las cuales Hel308 es un miembro (adaptado de Tuteja et al., "Unraveling DNA Helicases: Motif, Structure, Mechanism and Function", *European Journal of Biochemistry* 271 (10): 1849-1863 (2004)).
- Las figuras 17A-17D muestran señales ilustrativas generadas con Mbu helicasa Hel308, Tga helicasa Hel308 y eventos de translocación de la polimerasa phi29 utilizando ciertos parámetros, de acuerdo con algunas realizaciones.
- La Fig. 18 ilustra esquemáticamente las etapas en un método ejemplar para realizar ensayos utilizando translocación fraccional para caracterizar señuelos de polinucleótidos, de acuerdo con algunas realizaciones.
- La Fig. 19A ilustra esquemáticamente un aspecto de un modelo de Markov oculto (HMM) ejemplar usado para caracterizar una señal de translocación de un solo etapa de un polinucleótido a través de un poro.
- La Fig. 19B ilustra esquemáticamente un aspecto de un HMM ejemplar utilizado para caracterizar señales de la translocación en pasos fraccionales de un polinucleótido a través de un poro utilizando una helicasa Hel308, de acuerdo con algunas realizaciones.
- La Fig. 20A ilustra los resultados ejemplares de la secuenciación de novo utilizando etapas fraccionarias, de acuerdo con algunas realizaciones.
- Las figuras 20B-20C ilustran resultados ejemplares de la comparación de patrones utilizando etapas fraccionarias, de acuerdo con algunas realizaciones.
- Las figuras 21A-21C ilustran esquemáticamente señales que pueden generarse en función del tiempo para diferentes translocaciones de un polinucleótido a través de un poro, de acuerdo con algunas realizaciones.
- Las figuras 22A-22D ilustran pasos en métodos ilustrativos para usar la información proporcionada por la translocación fraccional de un polinucleótido a través de un poro, de acuerdo con algunas realizaciones.
- La Fig. 23 ilustra señales simuladas ejemplares que pueden generarse en función del tiempo para una primera secuencia polinucleotídica ilustrativa (SEC ID NO: 89) y una segunda secuencia polinucleotídica ilustrativa (SEC ID NO: 90) adecuada para su uso como señuelos respectivos, según algunas realizaciones.
- Las figuras 24A-24D ilustran señales simuladas ejemplares que pueden generarse en función del tiempo para la primera y segunda secuencias polinucleotídicas ilustrativas adecuadas para su uso como señuelos respectivos, de acuerdo con algunas realizaciones.
- Las figuras 25A y 25B ilustran respectivamente señales simuladas ejemplares que pueden generarse en función del tiempo para la primera y segunda secuencias polinucleotídicas ilustrativas adecuadas para su uso como señuelos respectivos, de acuerdo con algunas realizaciones.
- Las figuras 26A-26D ilustran respectivamente señales medidas ejemplares que se generaron en función del tiempo para la primera y la segunda secuencia polinucleotídica ilustrativa adecuadas para su uso como señuelos respectivos, de acuerdo con algunas realizaciones.
- Descripción detallada de las realizaciones
- La presente descripción proporciona métodos y composiciones para caracterizar un polinucleótido diana que incluye caracterizar la secuencia de un polinucleótido diana utilizando uno o más pasos de translocación fraccional de la translocación del polinucleótido diana a través de un poro.
- Al desarrollar la tecnología de secuenciación de nanoporos se puede lograr un cierto nivel de translocación controlada de un polinucleótido a través de un nanoporo bajo la guía de un motor molecular, tal como una helicasa, translocasa o polimerasa contra (por ejemplo, para resistir la fuerza generada por) una diferencia de potencial eléctrico. Los motores moleculares pueden mover el polinucleótido de manera gradual, normalmente con uno o más nucleótidos por etapa. Este trinquete controlado ralentiza la translocación de polinucleótidos a través del nanoporo de una tasa nativa de $\mu\text{sec/nucleótido}$ a msec/nucleótido .
- Los motores moleculares pueden utilizar la energía de la hidrólisis de nucleótidos para impulsar la translocación de los polinucleótidos a través del nanoporo. Una helicasa es un ejemplo en el que la hidrólisis de ATP es la fuente de energía para la translocación de polinucleótidos. La representación en la fig. 1 ilustra el modelo electrostático del

gusano medidor para la translocación de un polinucleótido a través de la helicasa (ver Frick et al., *Current Pharmaceutical Design*, 12: 1315-1338 (2006)). En este modelo, un polinucleótido de una sola hebra se mantiene en una hendidura con carga negativa que separa los dos dominios RecA de una helicasa de un tercer dominio. En ausencia de ATP, un residuo insignia (por ejemplo, Trp501 en helicasa HCV) y un residuo pinza (por ejemplo, Arg393 en helicasa HCV) evitan que el polinucleótido monocatenario se deslice a través de una hendidura. Tras la unión de ATP, los dominios RecA giran, moviendo la Arg-pinza cargada positivamente. La Arg-pinza atrae al polinucleótido monocatenario cargado negativamente, que a su vez salva al hito. El polinucleótido monocatenario es luego repelido por la hendidura cargada negativamente, y el polinucleótido monocatenario se transloca a través de la helicasa hasta que se hidroliza el ATP. Por lo tanto, en este modelo ejemplar, la translocación de polinucleótidos a través de una helicasa implica al menos dos etapas: una primera etapa donde la helicasa se une a ATP y experimenta un cambio conformacional, y una segunda etapa donde se hidroliza el ATP y el polinucleótido se transloca a través de la helicasa.

La Fig. 1B ilustra esquemáticamente una primera composición ejemplar que incluye un poro en contacto con una helicasa Hel308, de acuerdo con algunas realizaciones. En la Fig. 1B, el círculo semitransparente lleno de marcas indica una helicasa Hel308 tal como se proporciona en este documento. La línea recta indica un polinucleótido, y las líneas de puntos indican una longitud arbitraria del polinucleótido. La flecha gris grande indica la dirección del movimiento del polinucleótido hacia dentro o fuera del poro, y la flecha negra grande indica la dirección de la translocación de la helicasa a lo largo del polinucleótido, que va de 3' a 5'. En la realización ilustrada, el poro (objeto cónico en forma de embudo) se asienta en una membrana (líneas horizontales dobles), aunque se pueden usar adecuadamente otras configuraciones de poros. En la realización ilustrada en la fig. 1B, la dirección del movimiento del polinucleótido puede ser con el campo aplicado generado por una diferencia de potencial a través del poro (ilustrativamente una diferencia de potencial eléctrico de 180 V, aunque pueden usarse adecuadamente otras diferencias de potencial). Para hacer que la dirección del movimiento del polinucleótido sea nuevamente el campo aplicado generado por una diferencia de potencial a través del poro, la orientación del ADN se puede voltear, tal como se describe con mayor detalle a continuación con referencia a las Figs. 15A-15C. Como se proporciona con mayor detalle en el presente documento, las helicasas Hel308 pueden causar la translocación fraccional de un polinucleótido a través de un poro, lo que puede facilitar la caracterización del nucleótido. Por ejemplo, dicha translocación fraccional puede producir una o más señales, en función de las cuales se puede caracterizar el polinucleótido. La una o más señales pueden incluir una señal eléctrica tal como se describe en otra parte en este documento, o puede incluir una señal óptica tal como se describe en otra parte en este documento. Las señales eléctricas ejemplares pueden ser una medida seleccionada entre la medición de la corriente, el voltaje, el efecto túnel, la resistencia, el potencial, el voltaje, la conductancia y la medición eléctrica transversal.

De manera ilustrativa, a medida que la helicasa Hel308 transloca fraccionalmente el polinucleótido a través del poro, el paso de diferentes bases de nucleótidos dentro del poro puede causar cambios medibles en una corriente eléctrica a través del poro; dicha corriente eléctrica puede denominarse corriente de "bloqueo". Como se describe con mayor detalle en el presente documento, una o más características del polinucleótido, tal como una secuencia del polinucleótido, una modificación del polinucleótido, una longitud del polinucleótido, una identidad del polinucleótido, una fuente del polinucleótido, o una estructura secundaria del polinucleótido, o cualquier combinación adecuada de los mismos, puede determinarse en función de los cambios en la señal, por ejemplo, en función de los cambios en una corriente a través del poro, cuyos cambios se basan en pasos de translocación fraccional por la helicasa Hel308 del polinucleótido a través del poro. En las realizaciones en las que el poro es asimétrico, por ejemplo, incluye una boca de poro con un diámetro mayor que una base de poro (por ejemplo, tal como para MspA), la helicasa Hel308 puede estar en contacto con la boca del poro, como se ilustra en la FIG. 1B. Dicha configuración puede denominarse configuración "hacia adelante". De manera más general, una "configuración hacia adelante" puede referirse a la dirección en la que las moléculas pueden atravesar la naturaleza de los poros, independientemente de si el poro incluye una boca de poros más ancha que la base de los poros. Alternativamente, la "dirección hacia adelante" puede definirse arbitrariamente.

La Fig. 1C ilustra esquemáticamente las etapas en un método para caracterizar un polinucleótido diana. El método puede incluir una etapa de aplicar una diferencia de potencial a través de un poro en contacto con una helicasa Hel308 y un polinucleótido diana (etapa 110). De una manera análoga a la descrita más adelante con referencia a las Figs. 13A-13E y 14A-14D, la translocación del polinucleótido puede ser en una dirección opuesta a la fuerza aplicada causada por la diferencia de potencial en el polinucleótido que se transloca a través del poro, o la translocación del polinucleótido puede ser en una dirección con la fuerza aplicada causada por la diferencia de potencial en el polinucleótido que se traduce a través del poro. Opcionalmente, las etapas 110-130 se pueden repetir una o más veces. La etapa de translocación fraccional (etapa 120) puede incluir una primera etapa de translocación fraccional de un ciclo de translocación completo de la helicasa Hel308, o puede incluir una segunda etapa de translocación de un ciclo de translocación completo de la helicasa Hel308.

Como se usa en el presente documento, el término "polinucleótido" se refiere a un ácido desoxirribonucleico (ADN), ácido ribonucleico (ARN) o un análogo de estos. Un polinucleótido puede ser monocatenario, bicatenario o contener tanto una secuencia monocatenaria como bicatenaria. Las moléculas polinucleotídicas pueden originarse en la forma de ADN bicatenario (ADNds) (por ejemplo, ADN genómico, productos de la PCR y amplificación y similares), o pueden originarse en forma monocatenaria como ADN (ADNss) o ARN y se pueden convertir a la forma ADNds y viceversa. La secuencia precisa de una molécula de polinucleótido puede ser conocida o desconocida. Los siguientes son ejemplos ilustrativos de polinucleótidos: un gen o fragmento de gen (por ejemplo, una sonda, cebador, marcador EST

o SAGE), ADN genómico, fragmento de ADN genómico, exón, intrón, ARN mensajero (ARNm), ARN de transferencia, ARN ribosómico, ribozima, ADNc, polinucleótido recombinante, polinucleótido sintético, polinucleótido ramificado, plásmido, vector, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sonda de ácido nucleico, cebador o copia amplificada de cualquiera de los anteriores.

Un polinucleótido puede estar compuesto de nucleótidos o análogos de nucleótidos. Un nucleótido contiene típicamente un azúcar, una nucleobase y, al menos, un grupo fosfato. Un nucleótido puede ser abásico (es decir, carente de una nucleobase). Los nucleótidos incluyen desoxirribonucleótidos, desoxirribonucleótidos modificados, ribonucleótidos, ribonucleótidos modificados, nucleótidos peptídicos, nucleótidos peptídicos modificados, nucleótidos de esqueleto de azúcar fosfato modificado y mezclas de los mismos. Los ejemplos de nucleótidos incluyen, por ejemplo, monofosfato de adenosina (AMP), difosfato de adenosina (ADP), trifosfato de adenosina (ATP), monofosfato de timidina (TMP), difosfato de timidina (TDP), monofosfato de citidina (CMP), difosfato de citidina (CDP), trifosfato de citidina (CTP), monofosfato de guanosina (GMP), difosfato de guanosina (PIB), trifosfato de guanosina (GTP), monofosfato de uridina (UMP), difosfato de uridina (UMP), monofosfato de desoxiadenosina (dAMP), difosfato de desoxiadenosina (dADP), trifosfato de desoxiadenosina (dATP), monofosfato de deoxitimidina (dTTP), difosfato de deoxitimidina (dTDP), trifosfato de deoxitimidina (dTTP), difosfato de desoxicitidina (dCDP), trifosfato de desoxicitidina (dCTP), monofosfato de desoxiguanosina (dGMP), desoxiguanosina difosfato (dGDP), desoxiguanosina trifosfato (dGTP), desoxiuridina monofosfato (dUMP), desoxiuridina difosfato (dUDP) y desoxiuridina trifosfato (dUTP). Los análogos de nucleótidos que incluyen una nucleobase modificada también se pueden usar en los métodos descritos en este documento. Las nucleobases modificadas ejemplares que pueden incluirse en un polinucleótido, ya sea que tengan una estructura principal o estructura análoga nativa, incluyen, por ejemplo, inosina, xantina, hipoxantina, isocitosina, isoguanina, 2-aminopurina, 5-metilcitosina, 5-hidroximetil citosina, 2-aminoadenina, 6-metil adenina, 6-metil guanina, 2-propil guanina, 2-propil adenina, 2-tio-uracilo, 2-tiotimina, 2-tiocitosina, 15-halouracilo, 15-halocitosina, 5-propinil uracilo, 5-propinil citosina, 6-azouracilo, 6-azo citosina, 6-azo-timina, 5-uracilo, 4-tiouracilo, 8-halo adenina o guanina, 8-amino adenina o guanina, 8-tiol adenina o guanina, 8-tioalquil adenina o guanina, 8-hidroxil adenina o guanina, uracilo o citosina sustituida con 5-halo, 7-metilguanina, 7-metiladenina, 8-azaguanina, 8-azadenina, 7-deazaguanina, 7-deazadenina, 3-deazaguanina, 3-deazadenina o similares. Como se sabe en la técnica, ciertos análogos de nucleótidos no pueden incorporarse a un polinucleótido, por ejemplo, análogos de nucleótidos tales como adenosina 5'-fosfosulfato.

Como se usa en este documento, el término "poro" pretende significar una estructura que se extiende a través de una barrera, tal como una membrana, que permite que los iones y/o moléculas solubles en agua crucen de un lado de la barrera al otro lado de la barrera. Los poros pueden aparecer, pero no necesariamente, en una membrana. Por ejemplo, una barrera que normalmente inhibe el paso de iones o moléculas solubles en agua puede incluir una estructura de poros que se extiende a través de la barrera para permitir el paso de los iones o de las moléculas solubles en agua de un lado de la barrera al otro lado de la barrera. Los poros (por ejemplo, poros transmembrana) incluyen, por ejemplo, poros biológicos, poros en estado sólido y poros híbridos en estado biológico y sólido.

Como se usa en este documento, la expresión "poro biológico" pretende significar un poro, que está hecho de materiales de origen biológico, que se extienden a través de una barrera, incluyendo por ejemplo una membrana que permite que los iones y/o moléculas solubles en agua se crucen de un lado de la barrera al otro lado de la barrera. El origen biológico se refiere al material derivado o aislado de un entorno biológico, como un organismo o célula, o una versión fabricada sintéticamente de una estructura biológicamente disponible. Los poros biológicos incluyen, por ejemplo, poros de polipéptidos y poros de polinucleótidos.

Como se usa en el presente documento, la expresión "poro polipeptídico" pretende significar uno o más polipéptidos que se extienden a través de una barrera como una membrana, por ejemplo, y permite que los iones y/o moléculas solubles en agua fluyan de un lado de la barrera al otro lado de la barrera. Un poro polipeptídico puede ser un monómero, un homopolímero o un heteropolímero. Las estructuras de los poros de polipéptidos incluyen, por ejemplo, un poro de haz de hélice α y un poro en barril β , así como todos los demás bien conocidos en la técnica. Poros polipeptídicos ejemplares incluyen α -hemolisina, porina A de *Mycobacterium smegmatis*, gramicidina A, maltoporina, OmpF, OmpC, PhoE, Tsx, F-pilus, SP1 (Wang et al., Chem. Commun., 49: 1741-1743, 2013) y porina mitocondrial (VDAC)XX, Tom40, (Patente de Estados Unidos No. 6.015.714 y Derrington et al., Proc. Natl. Acad. Sci. USA, 107: 16060 (2010)). "Mycobacterium smegmatis porina A (MspA)" es una porina de membrana producida por *Mycobacteria*, que permite que las moléculas hidrófilas entren en la bacteria. MspA forma un octámero estrechamente interconectado y un barril beta transmembrana que se asemeja a una copa y contiene un canal/poro central.

Como se usa en el presente documento, la expresión "poro de polinucleótido" pretende significar uno o más polinucleótidos que se extienden a través de una barrera como una membrana, por ejemplo, y permite que los iones y/o moléculas solubles en agua fluyan de un lado de la barrera al otro lado de la barrera. Un poro de polinucleótido puede incluir, por ejemplo, un origami de polinucleótido.

Tal como se usa en el presente documento, la expresión "poro en estado sólido" pretende significar un poro que está hecho de materiales de origen no biológico, que se extienden a través de una barrera, tal como una membrana, por ejemplo, que permite a los iones y/o moléculas solubles en agua cruzar de un lado al otro lado de la barrera. Se entiende por estado sólido a los materiales que no son de origen biológico. Un poro en estado sólido puede ser de materiales orgánicos o inorgánicos. Los poros en estado sólido incluyen, por ejemplo, poros de nitruro de silicio, poros

de dióxido de silicio y poros de grafeno.

Tal como se usa en el presente documento, la expresión "poro híbrido en estado sólido y biológico" pretende significar un poro híbrido que está hecho de materiales de origen biológico y no biológico y que se extienden a través de una barrera, tal como una membrana, por ejemplo, permitiendo cruzar a iones hidratados y/o moléculas solubles en agua de un lado de la barrera al otro lado. Los materiales de origen biológico se definen anteriormente e incluyen, por ejemplo, polipéptidos y polinucleótidos. Un poro híbrido en estado sólido y biológico incluye, por ejemplo, un poro híbrido en estado sólido polipéptido y un poro en estado sólido polinucleótido.

Como se usa en este documento, el término "helicasa" pretende significar una proteína de unión a polinucleótido que tiene una actividad que utiliza la energía derivada de la hidrólisis de, por ejemplo, un trifosfato de nucleótido (NTP) para desenrollar los polinucleótidos de doble cadena. Desenrollar un polinucleótido de doble cadena da como resultado la translocación del polinucleótido a lo largo de su sitio activo. El término pretende incluir polipéptidos que tienen actividades que se translocan o se unen a polinucleótidos monocatenarios, así como a polinucleótidos parcialmente bicatenarios. Una "helicasa Hel308" es una helicasa de ADN dependiente de ATP y una superfamilia 2 helicasa. El miembro fundador, Mus308 de *Drosophila melanogaster*, consiste en un dominio de helicasa SF2 N-terminal fusionado a un dominio de polimerasa de ADN C-terminal. El Hel308 en *Homo sapiens* funciona como un SF2, una ADN helicasa de 3' a 5' con capacidad de procesamiento limitada. La helicasa Hel308 se usa indistintamente con la helicasa tipo ski2. Los homólogos útiles pueden consistir solo en un dominio de helicasa (es decir, estar ausente un dominio de polimerasa). Los homólogos de helicasa sólo están presentes en metazoos y arqueas. Los ejemplos de metazoos son Hel308 y Mus301 humanos. Ejemplos de arqueas son Tga y Mbu.

A menos que se describa explícitamente de otro modo en el presente documento, la expresión "sustrato de helicasa Hel308", como se usa en este documento, significa un nucleótido o análogo de nucleótido que puede ser hidrolizado por la helicasa y proporciona energía para desenrollar un polinucleótido de doble cadena o parcialmente de doble cadena o transloca un polinucleótido monocatenario. Un sustrato común para una helicasa Hel308 incluye ATP. Sin embargo, otros sustratos de helicasa Hel308 dentro del significado del término incluyen nucleótidos distintos de ATP como los descritos anteriormente y análogos de nucleótidos que pueden ser hidrolizados por una helicasa Hel308. Los análogos ejemplares incluyen, por ejemplo, análogos de fosfato tales como análogos de gamma-tiol, análogos de alfa-tiol y similares, ATPγS, ATPαS, AMP, PNP, ApCpp, AppCp y AppNHp.

Tal como se usa en el presente documento, el término "translocaciones" o "translocación" pretende significar el movimiento de un polinucleótido diana a lo largo (o dentro de) una helicasa y/o un poro.

Como se usa en el presente documento, la expresión "ciclo de translocación completo", cuando se usa en referencia a una helicasa, significa un intervalo completo para el movimiento de una unidad de uno o más nucleótidos de un polinucleótido diana a lo largo de la helicasa y/o el poro. El intervalo completo puede comenzar en cualquier punto del ciclo, y puede, por ejemplo, incluir el intervalo representado en la FIG. 3 que incluye las etapas de unión a ATP e hidrólisis de la ATP unida. Por consiguiente, un ciclo de translocación completo como se usa en este documento puede comenzar en la unión del sustrato de nucleótido y terminar en la hidrólisis del sustrato de nucleótido. De manera similar, un ciclo completo de translocación puede comenzar en la hidrólisis del sustrato del nucleótido y terminar en la unión del nucleótido. De manera similar, un ciclo completo de translocación puede comenzar en cualquier punto entre los dos puntos de inicio ejemplificados anteriormente, siempre y cuando concluya en la etapa justo antes del punto de partida.

Tal como se usa en el presente documento, la expresión "etapa de translocación fraccional", cuando se usa en referencia a una helicasa, significa un evento detectable que caracteriza una parte de un ciclo de translocación completo. Por ejemplo, una etapa de translocación fraccional puede ser una translocación parcial de una unidad de uno o más nucleótidos de un polinucleótido diana a lo largo de la helicasa y/o el poro. En realizaciones particulares, puede producirse una etapa fraccionaria entre la unión de ATP y la hidrólisis cuando se produce un cambio conformacional. El cambio conformacional divide efectivamente un ciclo completo de translocación en al menos dos pasos de translocación parcial o fraccional. Un etapa fraccional puede ser concomitante o no con el movimiento del ácido nucleico a lo largo de la helicasa.

Tal como se usa en este documento, el término "señal" pretende significar un indicador que representa información. Las señales incluyen, por ejemplo, una señal eléctrica y una señal óptica.

Tal como se usa en el presente documento, la expresión "señal eléctrica" pretende significar un indicador de una cualidad eléctrica que representa información. El indicador puede ser, por ejemplo, corriente, voltaje, efecto túnel, resistencia, potencial, voltaje, conductancia; medición eléctrica y transversal. Una "corriente electrónica" se refiere a un flujo de carga eléctrica. La carga eléctrica fluye cuando se aplica una diferencia de potencial eléctrico a través del poro.

Tal como se usa en el presente documento, la expresión "señal óptica" pretende significar un indicador de una cualidad óptica que representa información. Las señales ópticas incluyen, por ejemplo, señales de fluorescencia y señales Raman.

Como se usa en el presente documento, el término "homología" pretende significar una similitud de secuencia entre

dos polinucleótidos o entre dos polipéptidos. La similitud se puede determinar comparando una posición en cada secuencia, que se puede alinear con propósitos comparativos. Un grado de similitud entre secuencias es una función del número de posiciones coincidentes u homólogas compartidas por las secuencias. La alineación de dos secuencias para determinar su porcentaje de similitud de secuencia se puede realizar utilizando programas de software conocidos en la técnica, como, por ejemplo, los descritos en Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley and Sons, Baltimore, MD. (1999). Preferiblemente, los parámetros predeterminados se utilizan para la alineación, cuyos ejemplos se exponen a continuación. Un programa de alineación bien conocido en la técnica que se puede usar es el ajuste de BLAST en los parámetros predeterminados. En particular, los programas son BLASTN y BLASTP, utilizando los siguientes parámetros predeterminados: Código genético = estándar; filtro = ninguno; hebra = ambos; corte = 60; espera = 10; Matriz = BLOSUM62; Descripciones = 50 secuencias; ordenar por = ALTA PUNTUACIÓN; Bases de datos = no redundante, traducciones de GenBank + EMBL + DDBJ + PDB + GenBank CDS + SwissProtein + SPupdate + PIR. Los detalles de estos programas se pueden encontrar en el Centro Nacional de Información Biotecnológica.

La presente descripción proporciona un método para caracterizar un polinucleótido diana. El método incluye: (a) aplicar una diferencia de potencial a través de un poro en contacto con una helicasa Hel308 y un polinucleótido diana; (b) medir una o más señales producidas por uno o más pasos de translocación fraccional del polinucleótido diana a través del poro, y (c) caracterizar el polinucleótido diana a partir de la señal eléctrica de las etapas de translocación fraccional.

Como se describe en el presente documento, los polinucleótidos incluyen ácido desoxirribonucleico (ADN), ácido ribonucleico (ARN) o análogos de los mismos. Un polinucleótido generalmente contendrá enlaces fosfodiéster, aunque en algunos casos un polinucleótido también puede tener estructuras alternativas, que incluyen, por ejemplo, fosforamida (Beaucage, et al., *Tetrahedron*, 49 (10): 1925 (1993) y sus referencias; Letsinger, J. *Org. Chem.*, 35: 3800 (1970); Sprinzl, et al., *Eur. J. Biochem.*, 81: 579 (1977); Letsinger, et al., *Nucl. Acids Res.*, 14: 3487 (1986); Sawai, et al., *Chem. Lett.*, 805 (1984); Letsinger, et al., *J. Am. Chem. Soc.*, 110: 4470 (1988); y Pauwels, et al., *Chemica Scripta*, 26: 141 (1986)), fosforotioato (Mag, et al., *Nucleic Acids Res.*, 19: 1437 (1991) y Patente de Estados Unidos No. 5.644.048), fosforoditioato (Briu, et al., *J. Am. Chem. Soc.*, 111: 2321 (1989)), enlaces O-metilfosforamida (ver Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press), y estructuras y enlaces peptídicos de ácidos nucleicos (ver Egholm, *J. Am. Chem. Soc.*, 114: 1895 (1992); Meier, et al., *Chem. Int. Ed. Engl.*, 31: 1008 (1992); Nielsen, *Nature*, 365: 566 (1993); Carlsson, et al., *Nature*, 380: 207 (1996)). Otros polinucleótidos incluyen aquellos con estructuras positivas (Denpcy, et al., *Proc. Natl. Acad. Sci. USA*, 92: 6097 (1995)); estructuras no iónicas (Patente de EE.UU. N° 5.386.023, 5.637.684, 5.602.240, 5.216.141 y 4.469.863; Kiedrowshi, et al., *Angew. Chem. Int. Ed. English*, 30: 423 (1991); Letsinger, et al., *J. Am. Chem. Soc.*, 110: 4470 (1988); Letsinger, et al., *Nucleosides & Nucleotides*, 13: 1597 (1994); Capítulos 2 y 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui y P. Dan Cook; Mesmaeker, et al., *Bioorganic & Medicinal Chem. Lett.*, 4: 395 (1994); Jeffs, et al., *J. Biomolecular NMR*, 34:17 (1994); *Tetrahedron Lett.*, 37: 743 (1996)) y estructuras sin ribosa, que incluyen las descritas en las Patentes de EE.UU. Nos. 5.235.033 y 5.034.506, y los capítulos 6 y 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y. S. Sanghui y P. Dan Cook. Las moléculas de polinucleótido que contienen uno o más azúcares carbocíclicos también se incluyen dentro de la definición de polinucleótido (ver Jenkins, et al., *Chem. Soc. Rev.*, (1995) pp. 169-176). Varios polinucleótidos se describen en Rawls, C & E News, 2 de junio de 1997, página 35.

El polinucleótido diana puede caracterizarse de acuerdo con los métodos de la presente descripción. Los polinucleótidos ejemplares incluyen, por ejemplo, un gen o fragmento de gen (por ejemplo, una sonda, cebador, marcador EST o SAGE), ADN genómico, fragmento de ADN genómico, exón, intrón, ARN mensajero (ARNm), ARN de transferencia, ARN ribosómico, ribozima, ADNc, polinucleótido recombinante, polinucleótido sintético, polinucleótido ramificado, plásmido, vector, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sonda de ácido nucleico, cebador o copia amplificada de cualquiera de los anteriores.

Un polinucleótido diana usado en realizaciones particulares en el presente documento puede ser de cualquiera de una variedad de longitudes, siendo típicamente de longitud suficiente para extenderse a través de un poro y estar unido en un lado del poro por una helicasa. En general, tal longitud tiene al menos aproximadamente 10 nucleótidos de longitud. Sin embargo, numerosas longitudes más largas que este tamaño mínimo son aplicables para la caracterización utilizando los métodos de la presente descripción. Las longitudes ejemplares de un polinucleótido útil incluyen, por ejemplo, al menos aproximadamente 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 300, 400, 500, 1.000, 5.000 ó 10.000, 100.000 nucleótidos o más. De forma alternativa o adicional, la longitud no puede ser mayor de 1.000.000, 100.000, 10.000, 1.000, 100 nucleótidos o menos. Por consiguiente, un polinucleótido que puede secuenciarse usando los métodos de la presente descripción puede variar, por ejemplo, desde polinucleótidos cortos, fragmentos, ADNc, genes y fragmentos genómicos.

El polinucleótido utilizado en los métodos de la presente descripción puede ser monocatenario, bicatenario o contener tanto secuencias monocatenarias como bicatenarias. Las moléculas de polinucleótido pueden originarse en un polinucleótido de doble cadena (por ejemplo, dsDNA) y pueden convertirse en un polinucleótido de cadena simple. Las moléculas polinucleótídicas también pueden originarse en un polinucleótido monocatenario (por ejemplo, ssDNA, ssRNA), y el ssDNA puede convertirse en un polinucleótido bicatenario. En algunos aspectos de la presente descripción, el polinucleótido bicatenario o parcialmente bicatenario incluye un polinucleótido bloqueante. Dichas especies de polinucleótidos pueden incluir aquellas ejemplificadas en relación con las Figs. 13A-13E, 14A-14D y 15A-

15C en este documento. Modos ejemplares de translocación de polinucleótidos a través de un poro se exponen en el documento WO 2013/057495.

En algunos aspectos, la presente descripción proporciona un método para caracterizar un polinucleótido diana. El método incluye identificar: (1) la secuencia del polinucleótido diana; (2) la modificación del polinucleótido diana; (3) la longitud del polinucleótido diana; (4) la identidad del polinucleótido diana; (5) la fuente del polinucleótido diana, o (6) la estructura secundaria del polinucleótido diana.

La secuencia del polinucleótido se refiere a la estructura primaria del polinucleótido o al orden secuencial de los nucleótidos en una molécula de polinucleótido. La secuencia del polinucleótido se puede determinar caracterizando los nucleótidos en el polinucleótido diana utilizando las señales producidas por las etapas de translocación fraccional del polinucleótido diana a través del poro.

Una modificación del polinucleótido se refiere a cualquier modificación covalente o no covalente de un nucleótido en el polinucleótido, que incluye, por ejemplo, la metilación de nucleótidos o la hidroximetilación. De hecho, las modificaciones pueden incluir cualquier número de análogos de nucleótidos que pueden incorporarse en una cadena de polinucleótidos, incluyendo, por ejemplo, 8-oxoguanosina, 5-formilcitosina y 5-carboxilcitosina y otros descritos en otra parte en este documento. La modificación de un nucleótido proporciona un cambio correspondiente en la señal. Por consiguiente, se pueden determinar una o más modificaciones de un polinucleótido caracterizando los nucleótidos modificados en el polinucleótido diana y utilizando las señales producidas por las etapas de translocación fraccional del polinucleótido diana a través del poro.

La longitud del polinucleótido se refiere a los números de nucleótidos en el polinucleótido. La longitud del polinucleótido se puede determinar, por ejemplo, determinando la secuencia primaria del polinucleótido o midiendo su tiempo de permanencia en un poro o contando el número de nucleótidos que pasan a través del poro. En algunas realizaciones, el tiempo de permanencia corresponde a la duración del cambio transitorio de la corriente. Un cambio transitorio puede considerarse cualquier desviación en la corriente del poro, debido a la presencia de un polinucleótido. En algunas realizaciones, la desviación da como resultado una reducción de la magnitud de la corriente. Esta reducción generalmente puede ser a lo sumo del 95%, 90%, 80%, 60%, 50%, 40%, 30%, 20% o 10% o menos de la corriente de poro desbloqueada original. Alternativa o adicionalmente, la reducción puede ser al menos del 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% o 90% o más. En algunos casos, el polinucleótido puede resultar en un aumento de la magnitud de la corriente en relación con el poro no bloqueado. La relación entre la duración y la longitud del polinucleótido se puede describir mediante una función matemática reproducible que depende de la condición experimental utilizada. La función puede ser una función lineal o no lineal (por ejemplo, sigmoidal o exponencial) para un tipo dado de polinucleótido (por ejemplo, ADN o ARN).

La identidad del polinucleótido se refiere al tipo de polinucleótido. La identidad también puede referirse al nombre del polinucleótido como se conoce en la técnica. Por ejemplo, la identidad de un polinucleótido puede ser, por ejemplo, ADN, ARN, un polinucleótido bicatenario, un polinucleótido monocatenario y/o un polinucleótido parcialmente bicatenario. La identidad de un polinucleótido también puede incluir la determinación del producto génico o la función estructural del polinucleótido. Por ejemplo, el polinucleótido puede codificar un polipéptido o puede ser un polinucleótido estructural tal como el ARN ribosómico. La identidad de un polinucleótido se puede determinar a partir de la secuencia de nucleótidos de todo o parte del polinucleótido, la secuencia de un segundo polinucleótido que es complementaria a todo o parte del polinucleótido, la secuencia de un ARN que está codificado por todo o parte del polinucleótido o la secuencia de una proteína que está codificada por todo o parte del polinucleótido. En ejemplos particulares, un polinucleótido puede identificarse por una secuencia de "marcador" o "señuelo" que forma parte del polinucleótido. En tales ejemplos, la identidad del polinucleótido puede asignarse por un patrón de señal esperado del marcador o señuelo. La fuente del polinucleótido puede referirse a las especies de origen del polinucleótido o a un origen sintético. La identidad y la fuente del polinucleótido se pueden determinar alineando la secuencia del polinucleótido en la base de datos de las secuencias de polinucleótidos, utilizando programas conocidos en la técnica, por ejemplo, el BLASTN.

La estructura secundaria del polinucleótido se refiere al cotejo de bases intramoleculares de regiones de auto-complementariedad en una molécula de polinucleótido. Las estructuras secundarias ejemplares incluyen, por ejemplo, una doble hélice, horquilla, bucle, protuberancia, dúplex, unión, vástago, pseudoknot, triple hélice, H-ADN, cabeza de martillo y ribozima autoadhesiva. La estructura secundaria del polinucleótido se puede determinar, por ejemplo, midiendo su cambio correspondiente en el tiempo de permanencia en un poro o midiendo el cambio correspondiente en la señal producida por las etapas de translocación fraccional.

Un poro es una estructura que se extiende a través de una barrera incluyendo, por ejemplo, una membrana que permite que los iones y/o las moléculas solubles en agua crucen de un lado de la barrera al otro. Los poros pueden aparecer, aunque no necesariamente, en una membrana. Por ejemplo, una barrera que normalmente inhibe el paso de iones o moléculas solubles en agua puede incluir una estructura de poros que se extiende a través de la barrera para permitir el paso de los iones o las moléculas solubles en agua de un lado al otro lado de la barrera. Una membrana de la presente descripción puede ser, por ejemplo, una barrera no permeable o semipermeable que separa dos cámaras de líquido que pueden tener las mismas o diferentes composiciones. Se puede usar cualquier membrana de acuerdo con la presente descripción, siempre que la membrana pueda configurarse para incluir un poro

transmembrana y mantener una diferencia de potencial a través de la membrana. Diferencias potenciales adecuadas se describen a continuación.

Se puede usar una variedad de membranas bien conocidas en la técnica en las composiciones y métodos de la presente descripción. Tales membranas conocidas en la técnica incluyen una variedad de diferentes estructuras y composiciones. Por ejemplo, la membrana puede tener una estructura monocapa o multicapa, siempre que se pueda incorporar un poro para la caracterización de un polinucleótido. Una capa en la membrana se refiere al material no permeable o semipermeable que forma la barrera. Los ejemplos de membranas monocapa y multicapa se describen con más detalle a continuación.

El material formador de membrana puede ser de origen biológico o no biológico. Un material que es de origen biológico se refiere a un material derivado o aislado a partir de un entorno biológico, como un organismo o célula, o una versión fabricada sintéticamente de una estructura biológicamente disponible. Una membrana ejemplar que está hecha de material que es de origen biológico incluye una bicapa lipídica. Un material que no es de origen biológico también se denomina material en estado sólido y puede formar una membrana en estado sólido.

Las bicapas lipídicas adecuadas y los métodos para fabricar u obtener bicapas lipídicas son conocidos en la técnica y se describen, por ejemplo, en la publicación de patente de EE.UU. US 2010/0196203 y la publicación de patente PCT WO 2006/100484. Las bicapas lipídicas adecuadas incluyen, por ejemplo, una membrana de una célula, una membrana de un orgánulo, un liposoma, una bicapa lipídica plana y una bicapa lipídica soportada. Una bicapa lipídica puede formarse, por ejemplo, a partir de dos capas opuestas de fosfolípidos, que se disponen de tal manera que sus colas hidrófobas se enfrenten entre sí para formar un interior hidrófobo, mientras que los grupos principales hidrófilos de los lípidos miran hacia el ambiente acuoso a cada lado de la bicapa. Las bicapas lipídicas también pueden formarse, por ejemplo, por el método de Montal y Mueller (Proc. Natl Acad Sci. USA., 1972; 69: 3561-3566), en la que una monocapa lipídica se transporta en una interfaz de solución acuosa/aire más allá de cada lado de una abertura que es perpendicular a esa interfaz. Normalmente, el lípido se agrega a la superficie de una solución de electrolito acuosa disolviéndolo primero en un disolvente orgánico y luego permitiendo que una gota del solvente se evapore sobre la superficie de la solución acuosa a cada lado de la abertura. Una vez que el disolvente orgánico se ha evaporado, las interfaces solución/aire en ambos lados de la abertura se mueven físicamente hacia arriba y hacia abajo más allá de la abertura hasta que se forma una bicapa. Otros métodos comunes de formación de bicapas incluyen la inmersión de puntas, la pintura de bicapas y la fijación de parches de las bicapas de liposomas. Una variedad de otros métodos para obtener o generar bicapas lipídicas son conocidos en la técnica e igualmente aplicables para su uso en las composiciones y métodos de la presente descripción.

Las membranas en estado sólido son también conocidas en la técnica y se describen, por ejemplo, en la publicación de patente PCT WO 2000/079257. Como se describió anteriormente, la membrana en estado sólido está hecha de una o más capas de materiales que no son de origen biológico. La membrana en estado sólido puede ser una monocapa, tal como un recubrimiento o película sobre un sustrato de soporte, o un elemento independiente. La membrana en estado sólido también puede ser un compuesto de múltiples capas de materiales en una configuración tipo sándwich. No hay ninguna limitación específica para los materiales que se pueden usar de acuerdo con la presente descripción, siempre que la membrana en estado sólido resultante se pueda configurar para incluir un poro transmembrana y configurarlo con una diferencia de potencial a través de la membrana. Las membranas en estado sólido pueden estar hechas de materiales orgánicos e inorgánicos, incluyendo, por ejemplo, materiales microelectrónicos, materiales aislantes tales como Si_3N_4 , Al_2O_3 y SiO_2 , polímeros orgánicos e inorgánicos tales como poliamida, copolímeros tribloque (por ejemplo, copolímeros de tribloques PMOXA-PDMS-PMOXA ABA anfífilicos), plásticos como el Teflon® o elastómeros, como el caucho de silicona de curado por adición de dos componentes y vidrios. Además, la membrana en estado sólido se puede fabricar a partir de una monocapa de grafeno, que es una lámina atómicamente delgada de átomos de carbono densamente empaquetados en una red de panel de abeja bidimensional, una multicapa de grafeno o una o más capas de grafeno mezcladas con una o más capas de otros materiales en estado sólido (publicación de patente PCT WO 2013/016486). Una membrana en estado sólido que contiene grafeno puede incluir al menos una capa de grafeno que es un nanoribono de grafeno o nanogap de grafeno, que se puede usar como un sensor eléctrico para caracterizar el polinucleótido diana (consulte la publicación de patente PCT WO 2013/016486). La membrana en estado sólido puede fabricarse por métodos conocidos en la técnica. Por ejemplo, la membrana de grafeno se puede preparar mediante deposición química de vapor (CVD) o exfoliación con grafito (publicación de patente PCT WO 2013/016486).

Las composiciones y métodos de la presente descripción pueden emplear un poro que se encuentra en una barrera para la caracterización de un polinucleótido diana. Un poro puede estar hecho a partir de materiales que tienen origen biológico o no biológico. En consecuencia, los poros incluyen, por ejemplo, poros biológicos, poros en estado sólido y poros híbridos en estado sólido y biológico.

Los poros pueden tener una funcionalidad asociada que facilita la detección de la secuencia de nucleótidos en un polinucleótido. Por ejemplo, los poros pueden incluir una enzima como la helicasa u otra funcionalidad unida, asociada o localizada cerca del poro para controlar la velocidad a la que los polinucleótidos transitan a través del poro. Los poros pueden tener circuitos o sensores de detección asociados, incluyendo, por ejemplo, circuitos de pinzamiento de parche, circuitos de electrodos con efecto túnel o circuitos de medición de la conductancia transversal (como un nanoribbon de grafeno o un nanogap de grafeno). Los poros también pueden incluir sensores ópticos que detectan

marcadores que incluyen, por ejemplo, restos fluorescentes o restos generadores de señal Raman en el polinucleótido que determina la secuencia de nucleótidos basada en la interacción de un fragmento con el poro (por ejemplo, pasando el fragmento a través del poro).

En realizaciones particulares, el poro biológico, que incluye un poro polipeptídico y un poro polinucleótido, se puede usar en las composiciones y métodos de la presente descripción, siempre que el poro tenga una zona de constricción que permita el paso del polinucleótido a través de la barrera (por ejemplo, la membrana). Una zona de constricción es una ubicación en el lumen del poro donde el bloqueo por un analito (por ejemplo, un polinucleótido o nucleótido) afecta a una señal detectable producida por el poro. Los poros que tienen una variedad de longitudes de zona de constricción pueden emplearse en la composición y los métodos de la presente descripción, incluyendo, por ejemplo, longitudes de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 nucleótidos. Alternativa o adicionalmente, se pueden usar longitudes de, a lo sumo, aproximadamente 10, 9, 8, 7, 6, 5, 4, 3, 2 ó 1 nucleótidos. Sin embargo, la longitud de la zona de constricción puede afectar a la calidad de la señal. Por ejemplo, las zonas de constricción más cortas pueden dar como resultado una mejor resolución de la translocación de nucleótidos o la precisión de la reconstrucción. En una realización, el poro biológico tiene una zona de constricción de aproximadamente cinco nucleótidos o menos, los cinco o menos de cinco nucleótidos ubicados en la zona de constricción modulan la señal eléctrica, que tiene una mejor resolución de la translocación de nucleótidos que la señal eléctrica obtenida a partir de más de cinco nucleótidos. En algunos casos, la mejora de la señal a ruido no da como resultado una mejora de la precisión de la secuencia para una constricción menor de 2 nt. Esto puede ocurrir si ya no se pueden detectar los homopolímeros mayores que la constricción más pequeña y la falta de re-lectura reduce la precisión cuando se omiten los nucleótidos debido al movimiento estocástico de la enzima. Por consiguiente, los poros de polipéptidos adecuados y los poros de polinucleótidos que tienen una zona de constricción de cinco nucleótidos o menos se pueden usar de acuerdo con la presente descripción. Dadas las enseñanzas y la guía que se proporcionan en este documento, los expertos en la materia entenderán qué zona de restricción de longitud es aplicable para cada necesidad particular. Por ejemplo, los expertos en la técnica pueden emplear poros con zonas de constricción más cortas en aplicaciones que requieran resultados de mayor calidad.

Un poro biológico es un poro que está hecho de materiales de origen biológico que se extienden a través de una barrera (por ejemplo, una membrana) que permite que los iones y/o las moléculas solubles en agua se crucen de un lado al otro lado de la barrera. Al igual que con las membranas utilizadas en este documento, cuando se refiere a poros, el origen biológico se refiere a una estructura derivada o aislada de un entorno biológico, tal como un organismo o célula, o una versión fabricada sintéticamente de una estructura disponible biológicamente. Los materiales de origen biológico incluyen, por ejemplo, polipéptidos y polinucleótidos. Por consiguiente, los poros biológicos incluyen, por ejemplo, poros de polipéptidos y poros de polinucleótidos.

Se puede usar un poro polipeptídico reconstituido en una barrera (por ejemplo, una membrana) como una bicapa lipídica para la secuenciación de nanoporos. Existe una variedad de poros polipeptídicos que pueden usarse de acuerdo con la presente descripción, siempre que el(los) polipéptido(s) puedan formar una zona de constricción que permita el paso del polinucleótido a través de la barrera (por ejemplo, la membrana). Dependiendo del polipéptido(s) involucrado(s), el poro del polipéptido puede ser un monómero, un homopolímero o un heteropolímero. El poro del polipéptido puede incluir varias subunidades de repetición, tales como 7 u 8 subunidades. Por consiguiente, el poro polipeptídico puede ser, por ejemplo, un poro hexamérico, heptamérico u octamérico.

Los poros de polipéptidos incluyen, por ejemplo, un poro de haz de hélice α y un poro de barril β , así como todos los otros conocidos en la técnica. El poro del haz de hélices α incluye un poro formado por hélices α . Los poros adecuados del haz de hélices α incluyen, por ejemplo, proteínas de membrana interna y proteínas de membrana externa α , como la toxina WZA y ClyA. El poro del barril β incluye un poro que está formado por hebras β . Los poros de barril β adecuados incluyen, por ejemplo, toxinas β , tales como α -hemolisina, toxina de ántrax y leucocidinas, y proteínas de membrana externa/porinas de bacterias, como la porina de *Mycobacterium smegmatis* (Msp), incluyendo MspA, la porina F de la membrana externa (OmpF), la porina G de la membrana externa (OmpG), la fosfolipasa A de la membrana externa y la lipoproteína autotransportadora de *Neisseria* (NalP). Otros poros incluyen, por ejemplo, lisenina (ver, por ejemplo, el documento WO 2013 153359, o el homólogo de MspA de *Norcadia farcinica*).

Un polipéptido de α -hemolisina es un poro de polipéptido heptamérico que puede usarse en los métodos y composiciones de la presente descripción. Está compuesto por un vestíbulo de 3,6 nm conectado a un barril β de ~5 nm de longitud, que contiene una constricción de 1,4 nm que permite el paso del polinucleótido monocatenario pero no del polinucleótido bicatenario. El poro cilíndrico de α -hemolisina de ~5 nm de largo puede alojar hasta aproximadamente 10 nucleótidos a la vez. Los nucleótidos ubicados en este barril β modulan significativamente la corriente de poros y posteriormente diluyen la firma iónica específica de un solo nucleótido en la constricción de poros más estrecha de 1,4 nm, lo que reduce la resolución general de la translocación de nucleótidos en aplicaciones de secuenciación.

MspA es un poro polipeptídico octamérico que puede usarse en las composiciones y métodos de la presente descripción. Contiene una única constricción de diámetro de ~1,2 nm con una longitud de constricción de 0,5 nm; el poro interno se conforma en un embudo cónico, en oposición a la estructura cilíndrica de la α -hemolisina. Derrington et al. demostró la capacidad de la MspA modificada por ingeniería genética para discriminar entre grupos de tres nucleótidos (AAA, GGG, TTT, CCC) con una mejora impresionante de 3,5 veces en la eficiencia de separación de nucleótidos respecto a la α -hemolisina nativa (Derrington et al., Proc. Natl Acad Sci. USA, 107: 16060 (2010)). Se

informó que en experimentos con polinucleótidos monocatenarios inmovilizados, se observó que tan solo tres nucleótidos dentro o cerca de la constricción de MspA contribuían a la corriente de poro, una mejora significativa respecto a los ~10 nucleótidos que se sabe modulan la corriente iónica en la α -hemolisina nativa. Los autores plantean la hipótesis de que esto podría mejorarse aún más hasta tal vez un solo nucleótido mediante mutagénesis específica del sitio, una diana de los futuros mutantes de MspA.

En algunos aspectos, el poro polipeptídico es una porina A (MspA) de *Mycobacterium smegmatis*. En algunos aspectos, la MspA tiene una secuencia de aminoácidos de SEC ID NO: 1 o tiene al menos el 15%, al menos el 20%, al menos el 25%, al menos el 30%, al menos el 35%, al menos el 40%, al menos el 45%, al menos el 50%, al menos el 55%, al menos el 60%, al menos el 65%, o al menos el 70%, al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, o al menos el 99% de homología con la SEC ID NO: 1.

MspA es un poro polipeptídico adecuado. Además, se pueden usar mutantes MspA en las composiciones y métodos de la presente descripción para regular la translocación de polinucleótidos a través del poro. El poro MspA utilizado en realizaciones en el presente documento puede tener la secuencia de aminoácidos de SEC ID NO: 1, correspondiente

GLDNELSLVDGQDRTLTVQQWDTFLNGVFPLDRNRLTREWFHSGRAKYIVAGPGADEFEGTLELGYQIGFPWSLGV
GINFSYTTNPILINNGNITAPPFGLNSVITPNLFPGVVISARLGNPGIQEVATFSVRVSGAKGGVAVSNAHGTVTGAAG
GVLLRPFARLIATGDSVTTYGEPWNMN, que es la secuencia de MspA con las siguientes mutaciones: D90N, D91N,
D93N, D118R, D134R y E139K. El mutante de poro MspA de SEC ID NO: 1 se denomina "M2 NNN". Se pueden usar
otros mutantes MspA en las composiciones y métodos de la presente descripción, que tienen al menos el 15%, al
menos el 20%, al menos el 25%, al menos el 30%, al menos el 35%, al menos el 40%, al menos 45%, al menos 50%,
al menos 55%, al menos 60%, al menos 65% o al menos 70%, al menos 75%, al menos 80%, al menos 85%, al menos
90%, al menos el 95%, o al menos el 99% de homología con la SEC ID NO: 1. Que una región de polipéptido o un
polipéptido (o una región de polinucleótido o un polinucleótido) tenga un cierto porcentaje (por ejemplo, 50%) de
homología con otra secuencia significa que, cuando están alineadas, ese porcentaje de aminoácidos (o bases de
nucleótidos) sea el mismo al comparar las dos secuencias. La alineación de dos secuencias para determinar su
porcentaje de identidad de secuencia se puede realizar utilizando programas de software conocidos en la técnica,
como se describe en este documento. Se pueden hacer mutaciones al polipéptido MspA nativo, incluidas las
inserciones, deleciones, sustituciones u otras modificaciones seleccionadas de regiones particulares o residuos de
aminoácidos específicos, de acuerdo con métodos que son conocidos en la técnica, incluida la mutagénesis específica
del sitio del ácido nucleico que codifica el polipéptido MspA (Zoller, MJ, Curr. Opin. Biotechnol., 3: 348-354, (1992)).
También se muestran mutantes MspA útiles en el documento US 2012/0055792A1.

Un polipéptido MspA nativo o mutante utilizado en las composiciones y métodos de la presente descripción se puede aislar mediante una variedad de métodos conocidos en la técnica, por ejemplo, sistemas de expresión recombinante, precipitación, filtración en gel, intercambio iónico, cromatografía de fase inversa y afinidad, y similares. Otros métodos conocidos se describen en Deutscher et al., Guide to Protein Purification: Methods in Enzymology, vol. 182, (Academic Press, (1990)). Alternativamente, el polipéptido MspA aislado o mutante aislado de la presente descripción se puede obtener usando métodos recombinantes conocidos. Los expertos en la técnica pueden elegir los métodos y condiciones para la purificación bioquímica del polipéptido MspA nativo o mutante de la presente descripción y la purificación se puede monitorear, por ejemplo, mediante un ensayo funcional.

Un método ejemplar para preparar un polipéptido MspA nativo o mutante es expresar el polinucleótido que codifica el polipéptido en una célula huésped adecuada, tal como una célula bacteriana, una célula de levadura u otra célula adecuada, usando métodos conocidos en la técnica y recuperando el polipéptido MspA nativo o mutante expresado, nuevamente utilizando métodos de purificación conocidos, como los descritos en este documento. Un polipéptido MspA nativo o mutante puede aislarse directamente de células que se han transformado con vectores de expresión como se describe en el presente documento. El polipéptido MspA nativo o mutante expresado de forma recombinante también puede expresarse como polipéptidos de fusión con marcadores de afinidad apropiados, como la glutatión S transferasa (GST) o poli His, y purificarse por afinidad. Un polipéptido MspA nativo o mutante también puede producirse por síntesis química usando cualquier método de síntesis de polipéptidos conocido por los expertos en la técnica.

También se puede usar para la secuenciación de nanoporos poros de polinucleótido reconstituidos en una barrera (por ejemplo, una membrana), como en una bicapa lipídica. El poro del polinucleótido es uno o más polinucleótidos que se extienden a través de una barrera (por ejemplo, una membrana) y permite que los iones y/o moléculas solubles en agua fluyan desde un lado al otro lado de la barrera. Se puede usar cualquier poro polinucleotídico de acuerdo con realizaciones particulares de la presente descripción, siempre que el polinucleótido o los polinucleótidos puedan formar una zona de constricción que permita el paso del polipéptido diana a través de la barrera (por ejemplo, la membrana). Los poros de polinucleótidos ejemplares incluyen, por ejemplo, un poro de origami polinucleotídico. El poro de origami polinucleotídico cuyos patrones se extienden en dos o tres dimensiones se puede hacer usando el "origami" como se describe en Rothmund, Nature, 440: 297-302 (2006). El origami es una técnica genérica que utiliza una hebra larga de polinucleótidos genómicos y muchas hebras de polinucleótidos "cortas" sintéticas más cortas para crear una estructura extendida. Las estructuras originales de origami eran estructuras sustancialmente bidimensionales. Desde entonces, la técnica del origami se ha extendido a estructuras tridimensionales (Douglas et al., Nature 459: 414-418 (2009); Ke et al., Nano Letters, 6: 2445-2447 (2009); Andersen et al., Nature 459: 73-76 (2009)).

También se puede usar un poro en estado sólido en las composiciones y métodos de la presente descripción. El poro en estado sólido es un poro que está hecho de materiales de origen no biológico que se extienden a través de una barrera (por ejemplo, una membrana) que permite que los iones y/o moléculas solubles en agua crucen de un lado al otro lado de la barrera.

- 5 Se puede formar un poro en estado sólido creando un poro en la barrera en estado sólido (por ejemplo, la membrana). Por consiguiente, de manera similar a la membrana en estado sólido y como se describe en el presente documento, el poro en estado sólido puede formarse mediante una variedad de materiales que abarcan tanto materiales inorgánicos como orgánicos.

- 10 Los poros en estado sólido adecuados incluyen, por ejemplo, óxido de aluminio, óxido de tantalio, óxido de titanio, dióxido de silicio, óxido de hafnio, óxido de circonio, nitruro de boro, nitruro de silicio, grafeno o nanolaminatos de los mismos (por ejemplo, grafeno- Al_2O_3), o cualquier combinación de los mismos (Publicación de patente PCT WO 2013016486A1). El poro en estado sólido se puede fabricar utilizando una herramienta de constitución de haces de iones controlada por retroalimentación personalizada o utilizando un haz de electrones convergente enfocado desde una pistola de emisión de campo (FEG) TEM para pulverizar como en una descomposición un nanoporo en la membrana, o cualquier otro método conocido en la técnica (publicación de patente PCT WO 2013016486A1). Por 15 ejemplo, un poro de nanolaminato de grafeno, tal como un poro de grafeno- Al_2O_3 , se puede hacer al perforar a través de una membrana de grafeno- Al_2O_3 usando un haz de electrones convergente enfocado a partir de un FEG TEM (Venkatesan et al., ACS Nano., 6: 441-450 (2012)).

- 20 Se puede usar un poro híbrido biológico y en estado sólido en las composiciones y métodos de la presente descripción. El poro híbrido en estado sólido y biológico es un poro híbrido, que está hecho de materiales de origen biológico y no biológico, que se extienden a través de una barrera (por ejemplo, una membrana) que permite que los iones y/o moléculas solubles en agua se crucen desde un lado al otro lado de la barrera. Los materiales de origen biológico se han definido anteriormente e incluyen, por ejemplo, polipéptidos y polinucleótidos. Los materiales de origen no biológico se denominan materiales en estado sólido, como se describe en este documento.

- 25 Por consiguiente, los poros híbridos en estado sólido y biológico incluyen, por ejemplo, un poro híbrido en estado sólido polipéptido y un poro híbrido en estado sólido polinucleótido. El poro híbrido polipéptido-estado sólido incluye uno o más polipéptidos y material en estado sólido. El poro híbrido de polinucleótido-estado sólido incluye uno o más polinucleótidos y material en estado sólido. Los poros híbridos biológicos y en estado sólido se fabrican mediante modificaciones de poros de polipéptidos o polinucleótidos, estando el poro en estado sólido (ver la publicación de patente PCT WO 2013/016486). Los ejemplos de poros de polipéptidos adecuados, poros de polinucleótidos y poros 30 en estado sólido se describen anteriormente.

- Un aparato de secuenciación de nanoporos puede tener poros singulares o múltiples. Se pueden usar múltiples poros como una matriz de nanoporos para caracterizar más de un nucleótido diana, que tienen composiciones iguales o diferentes. Los números ejemplares de los múltiples poros utilizados en este documento incluyen, por ejemplo, al 35 menos 1, 4, 16, 64, 256, 512, 1028, 4096, 16384, 32768, 100000, 1 millón, 10 millones de poros o más. En realizaciones preferidas, el número de poros múltiples será mayor que 4096. Las matrices de nanoporos son conocidas en la técnica y se describen, por ejemplo, en la publicación de patente PCT WO 2013/016486. Por ejemplo, se pueden fabricar matrices de alta densidad de poros en estado sólido de ~15 nm de diámetro utilizando litografía por haz de electrones y etapas de grabado químico con iones reactivos en membranas de $\text{SiN}/\text{Al}_2\text{O}_3$, facilitando el análisis a alto 40 rendimiento de las moléculas de polinucleótidos.

- Un método de la presente descripción puede utilizar una diferencia de potencial a través de una barrera (por ejemplo, una membrana). La diferencia de potencial puede ser una diferencia de potencial eléctrico, una diferencia de potencial químico o una diferencia de potencial electroquímico. Se puede imponer una diferencia de potencial eléctrico a través de la barrera (por ejemplo, la membrana) a través de una fuente de voltaje que inyecte o administre corriente a al 45 menos uno de los depósitos de líquido. Se puede imponer un potencial químico a través de la barrera por medio de una diferencia en la composición iónica de los dos grupos. Una diferencia de potencial electroquímico se puede establecer por una diferencia en la composición iónica de los dos grupos en combinación con un potencial eléctrico. Las diferentes composiciones iónicas pueden ser, por ejemplo, diferentes iones en cada grupo o diferentes concentraciones de los mismos iones en cada grupo.

- 50 La aplicación de un potencial eléctrico a través de un poro para forzar la translocación de un polinucleótido a través del poro es conocida en la técnica y se puede usar de acuerdo con la presente descripción (Deamer et al., Trends Biotechnol., 18: 147-151 (2000); Deamer et al., Ace Chem Res., 35: 817-825 (2002); y Li et al., Nat Mater., 2 (9): 611-615 (2003)). Un método de la presente descripción se puede llevar a cabo con un voltaje aplicado a través de un poro. El intervalo para el voltaje se puede seleccionar desde 40 mV hasta más de 1 V. Típicamente, un método de la presente descripción se ejecutará en el intervalo de 100 a 200 mV. En casos específicos, el método se realiza a 140 mV o 180 55 mV. No se requiere que los voltajes sean estáticos durante el movimiento del motor. La polaridad de voltaje se aplica típicamente de manera que el polinucleótido cargado negativamente se conduzca electroforéticamente en el poro. En algunos casos, el voltaje se puede reducir, o invertir la polaridad, para facilitar la función apropiada del motor.

En algunos casos, puede utilizarse la aplicación de diferenciales de presión para forzar la translocación de un

polinucleótido a través de un poro. Se pueden usar diferenciales de presión en lugar de potenciales eléctricos u otras diferencias potenciales en los métodos ejemplificados en este documento.

Los métodos de la presente descripción producen una o más señales que corresponden a la translocación de uno o más nucleótidos a través de un poro. Por consiguiente, como un polinucleótido diana transita a través de un poro, la corriente a través de la barrera cambia debido a un bloqueo de la constricción dependiente de la base, por ejemplo. La señal de ese cambio en la corriente se puede medir utilizando cualquiera de una variedad de métodos como se describe en el presente documento o como se conoce en la técnica. Cada señal es única para la especie de nucleótido(s) en el poro, de manera que la señal resultante se puede usar para determinar una característica del polinucleótido como se describió anteriormente. Por ejemplo, se puede determinar la identidad de una o más especies de nucleótidos que producen una señal característica. Las señales útiles en los métodos de la presente descripción incluyen, por ejemplo, señales eléctricas y señales ópticas, que se describen con más detalle a continuación. En algunos aspectos, la señal eléctrica puede ser una medida de corriente, voltaje, efecto túnel, resistencia, voltaje, conductancia; o medición eléctrica transversal (publicación de patente PCT WO 2013/016486). En algunos aspectos, la señal eléctrica es una corriente eléctrica que pasa a través de un poro.

Una señal eléctrica detectada en un método expuesto en este documento puede ser una corriente eléctrica, que es un flujo de carga eléctrica, que pasa a través de un poro (Deamer et al., *Trends Biotechnol.*, 18: 147-151 (2000); Deamer et al., *Ace Chem Res.*, 35: 817-825 (2002) y Li et al., *Nat Mater.*, 2 (9): 611-615 (2003)). Como se describe en el presente documento, las señales eléctricas se pueden medir utilizando el circuito de detección acoplado a un poro, por ejemplo, un circuito de pinzamiento de parche o un circuito de electrodo de efecto túnel. Los ejemplos de señales de voltaje, efecto túnel, resistencia y conductancia que se pueden detectar y los aparatos para su detección se conocen en la técnica tal como se describe, por ejemplo, en Wanunu, *Phys Life Rev.*, 9 (2): 125-58 (2012); y Venkatesan et al., *Nat Nanotechnol.*, 6 (10): 615-24 (2011).

Las señales ópticas útiles en los métodos de la presente descripción incluyen, por ejemplo, fluorescencia y señal Raman. Las señales ópticas pueden generarse acoplando el nucleótido diana con un marcador de generación de señal óptica, por ejemplo, un resto fluorescente o un resto generador de señal Raman. Por ejemplo, en dela Torre et al., *Nanotechnology*, 23 (38): 385308 (2012), se empleó el esquema óptico de microscopía de fluorescencia de reflexión interna total (TIRF, por sus siglas en inglés) para iluminar una amplia área de la membrana recubierta con TiO₂. En Soni et al., *Rev Sci Instrum.*, 81 (1): 014301 (2010), se usó un método para integrar dos modalidades de medición de una sola molécula, a saber, microscopía de reflexión interna total y detección eléctrica de biomoléculas utilizando nanoporos.

Como se describe en este documento, los poros se pueden acoplar con un circuito de detección, que incluye, por ejemplo, un circuito de pinzamiento de parche, un circuito de electrodo de efecto túnel o un circuito de medición de conductancia transversal (como un nanoribbon de grafeno o un nanogap de grafeno), para registrar las señales eléctricas en las presentes realizaciones. Además, el poro también puede acoplarse con un sensor óptico que detecta marcadores, por ejemplo, un resto fluorescente o un resto generador de señal Raman, en los polinucleótidos.

Los métodos de secuenciación de nanoporos pueden emplear un mecanismo para ralentizar la translocación de un polinucleótido diana a través de un poro. Por ejemplo, una proteína de unión a polinucleótido, como una helicasa, translocasa o polimerasa, puede unirse o incorporarse para regular la velocidad de translocación. La unión puede ser, por ejemplo, transitoria o persistente y puede estar mediada por el polinucleótido diana a medida que se extrae a través del poro o por una variedad de polipéptidos, enlazadores químicos o restos de captura conocidos en la técnica. Las técnicas ejemplares se describen en Manrao et al., *Nat Biotechnol.*, 30 (4): 349-353 (2012) y Cherf et al., *Nat Biotechnol.*, 30 (4): 344-348 (2102). Se puede usar una helicasa u otro motor molecular para ralentizar o detener la translocación de un polinucleótido diana a través de un poro. Por ejemplo, cuando se usa un motor que hidroliza los nucleótidos para efectuar la translocación, el nucleótido se puede omitir del motor y/o el motor se puede someter a un inhibidor (por ejemplo, un análogo de nucleótido no hidrolizable) de modo que el polinucleótido diana permanezca unido al motor y no se transloque apreciablemente a través del poro. En algunas realizaciones, posteriormente se puede permitir que se produzca la translocación suministrando nucleótidos al motor y/o eliminando el inhibidor. Un método de la presente descripción puede incluir una etapa de poner en contacto un poro con un polinucleótido diana y una helicasa Hel308 para controlar la tasa de translocación del polinucleótido a través del poro. Como se describe más adelante, una helicasa Hel308 puede caracterizarse como una helicasa de ADN dependiente de ATP y una helicasa de superfamilia 2. Dadas las enseñanzas y la guía proporcionadas en el presente documento, cualquier experto en la técnica podría seleccionar o adaptar adecuadamente cualquier helicasa Hel308 para su uso de acuerdo con las presentes realizaciones. Helicasas Hel308 adecuadas se describen con más detalle a continuación.

En algunos aspectos de un método expuesto en el presente documento, la translocación del polinucleótido diana es en una dirección opuesta a la dirección de una corriente a través del poro. En otros aspectos, una translocación del polinucleótido diana es en la misma dirección que la dirección de una corriente que pasa a través del poro.

Por consiguiente, un método de la presente descripción puede llevarse a cabo en al menos dos modos, en los que una translocación del polinucleótido diana es opuesta o está en la misma dirección, por ejemplo, que una corriente u otro potencial a través de un poro. Este resultado se puede lograr uniendo una helicasa Hel308 de la presente descripción a cualquiera de los extremos 5' o 3' del polinucleótido diana. Cuando se refiere a un polinucleótido de

doble cadena, la orientación 5' o 3' se refiere a una sola cadena dentro del polinucleótido de doble cadena. Por lo tanto, una helicasa Hel308 puede sacar o alimentar el polinucleótido desde o hacia dentro del poro, es decir, en la dirección contra la fuerza respecto al polinucleótido generado por el gradiente de voltaje (consulte las Figuras 13A-13E y 14A-14D), o utilizando la helicasa para regular la velocidad de translocación a medida que el polinucleótido se mueve en la misma dirección que la fuerza generada por el gradiente de voltaje u otro potencial (consulte las Figuras 15A-15C).

Las figuras 13A-13E ejemplifican la translocación de polinucleótidos controlada por helicasa Hel308 contra la fuerza generada por un potencial tal como un gradiente de voltaje, por ejemplo, basado en un complejo de polinucleótidos ternario con un punto de unión saliente 3' de una helicasa Hel308, y un ancla de colesterol bicapa, según algunas realizaciones. Círculo relleno (●) indica un fosfato de 5'. El diamante relleno (◆) indica un colesterol 3'. Un círculo semitransparente lleno de marcas indica la helicasa Hel308. Las líneas de puntos indican una longitud arbitraria. Las flechas grises grandes indican la dirección del movimiento del polinucleótido (con o contra el campo aplicado) del polinucleótido dentro o fuera del poro. Las flechas negras grandes indican la dirección de la translocación de la helicasa a lo largo del polinucleótido, que va de 3' a 5'. El poro (objeto cónico en forma de embudo) se asienta en la membrana (líneas horizontales dobles).

Las figuras 13A-13E ilustran el uso de un complejo de polinucleótido ternario con un sitio de unión saliente de 3' de helicasa Hel308 y el anclaje de la bicapa de colesterol para la secuenciación de polinucleótidos. El polinucleótido marcado con colesterol "ii" es opcional y se usa para hibridar con el polinucleótido diana "ii" que se translocará a través del poro y facilitará el reclutamiento de todo el complejo a la bicapa lipídica (Figura 13A). El fosfato 5' se atraviesa a través del poro, por ejemplo, por un gradiente de voltaje, lo que da como resultado que el extremo 5' del polinucleótido diana "ii" ingrese primero en el poro, y haga que el polinucleótido marcado con colesterol sea eliminado (FIG. 13B). A medida que el polinucleótido que contiene fosfato se empuja a través del poro hacia el lado trans, el segundo polinucleótido hibridado "iii" se elimina porque el poro es demasiado estrecho para permitir que el polinucleótido bicatenario se traslade (Figura 13C). Un propósito del polinucleótido iii es crear un sitio de unión de helicasa Hel308, generalmente un saliente de polinucleótido monocatenario 3' de aproximadamente 8 nucleótidos, al que puede unirse preferentemente una helicasa Hel308. Además, al obligar a la molécula de helicasa Hel308 a unirse al extremo 3' del polinucleótido de translocación, se maximiza la longitud del polinucleótido que se transloca a través del poro. El polinucleótido "iii" del complejo puede ser de cualquier longitud, incluida una longitud arbitraria, y el extremo 3' no necesita estar adyacente al extremo 5' del polinucleótido "i". Al llegar a la boca del poro, la helicasa Hel308 tira del polinucleótido contra el gradiente de voltaje a través de su actividad translocasa de 3' a 5' de vuelta a la cámara cis (Figuras 13D y 13E).

Las figuras 14A-14D también ilustran la translocación de polinucleótidos controlada por helicasa Hel308 contra la fuerza generada por un potencial tal como un gradiente de voltaje, por ejemplo, basado en un complejo de polinucleótidos ternario con un saliente de 3' de helicasa Hel308 y un ancla de dos capas de colesterol, según algunas realizaciones. Círculo relleno (●) indica un fosfato de 5'. El diamante relleno (◆) indica un colesterol 3'. Un círculo semitransparente lleno de marcas indica la helicasa Hel308. Las líneas de puntos indican longitud arbitraria. Las flechas grises grandes indican la dirección del movimiento del polinucleótido (con o contra el campo aplicado) del polinucleótido dentro o fuera del poro. Las flechas negras grandes indican la dirección de la translocación de la helicasa a lo largo del polinucleótido, que va de 3' a 5'. El poro (objeto cónico en forma de embudo) se asienta en la membrana (líneas horizontales dobles). Sin embargo, este esquema ejemplifica el uso de un único polinucleótido de hibridación "i" para crear un saliente de 3' en el polinucleótido diana "ii" para que una helicasa Hel308 se una a, y también contenga, un resto de colesterol opcional. Es posible que la helicasa Hel308 se pueda unir en cualquier lugar de las regiones monocatenarias del polinucleótido "ii". Se muestran múltiples moléculas de helicasa Hel308 y se indican con "E1", "E2" y "E3". La helicasa que llegue primero a la boca del poro iniciará el proceso de translocación controlada hacia el lado cis. En caso de fallo, se producirá una translocación no controlada hasta que la siguiente molécula de helicasa Hel308 unida llegue a la boca del poro y comience la translocación controlada.

Las figuras 15A-15C ejemplifican el uso de una helicasa Hel308 para regular la velocidad de translocación de los polinucleótidos a medida que el polinucleótido se mueve en la misma dirección que la fuerza generada por un potencial, como un gradiente de voltaje. Un círculo semitransparente lleno de marcas indica la helicasa Hel308. Las líneas de puntos indican longitud arbitraria. Las flechas grises grandes indican la dirección del movimiento del polinucleótido con el campo aplicado en el poro. Las flechas negras grandes indican la dirección de la translocación de la helicasa a lo largo del polinucleótido, que va de 3' a 5'. El poro (objeto cónico en forma de embudo) se asienta en la membrana (líneas horizontales dobles). En este esquema ejemplar, el polinucleótido diana entra primero en el extremo 3' del poro. La helicasa Hel308 controla la velocidad de translocación del polinucleótido al poro a medida que se traslada de 3' a 5' a lo largo del polinucleótido de translocación.

Como se describió anteriormente, una etapa de translocación fraccional en el contexto de una helicasa Hel308 puede referirse a una translocación parcial de uno o más nucleótidos del polinucleótido diana a lo largo de la helicasa y/o poro. En consecuencia, una etapa de translocación fraccional se refiere a una parte de una etapa de nucleótidos que es menor que el ciclo de translocación completo. Una etapa de translocación fraccional puede ocurrir entre la unión de ATP y la hidrólisis cuando se produce un cambio conformacional. Se pueden requerir uno o más pasos de translocación fraccional para una etapa completa de nucleótidos. El cambio conformacional divide efectivamente un ciclo completo de translocación en al menos dos etapas de translocación parcial o fraccional.

Las etapas de translocación parcial o fraccional pueden emplearse de la misma manera para generar una señal única para caracterizar uno o más nucleótidos que transiten a través de un poro. Por lo tanto, los métodos de la presente descripción pueden producir al menos dos señales eléctricas debido a un cambio en la corriente correspondiente a cada etapa de translocación fraccional para cada una o más translocación de nucleótidos a través de un poro. Por consiguiente, en algunos aspectos, la etapa de translocación fraccional incluye una primera etapa de translocación fraccional de un ciclo de translocación completo de la helicasa Hel308. En otros aspectos, la etapa de translocación fraccional incluye una segunda etapa de translocación fraccional de un ciclo de translocación completo de la helicasa Hel308. Cada primera o segunda etapa de translocación fraccional se puede usar sola o junto con su homólogo, por ejemplo, la segunda o primera etapa de translocación fraccional, respectivamente, para caracterizar uno o más nucleótidos que transitan una zona de constricción de un poro.

Por ejemplo, como se describe con más detalle en el Ejemplo I, una helicasa Hel308 puede unirse a ATP y sufrir un cambio conformacional, lo que proporciona un primer etapa de translocación fraccional, y la helicasa Hel308 puede translocar uno o más nucleótidos del polinucleótido diana a lo largo de la helicasa y/o el poro por hidrólisis de ATP, proporcionando una segunda etapa de translocación fraccional. Cualquiera de las etapas de translocación fraccional primera y segunda, o ambas, se pueden usar para determinar, por ejemplo, la secuencia de nucleótidos o los nucleótidos de uno o más nucleótidos que generan la señal. Cuando una señal es generada por más de un nucleótido, la parte del polinucleótido que genera la señal se denomina palabra. Por consiguiente, tales palabras de nucleótidos pueden tener una longitud de al menos 4, 5, 6, 7, 8, 9, 10 o más nucleótidos y corresponden a la longitud de la zona de constricción del poro. Alternativa o adicionalmente, las palabras de nucleótidos pueden tener una longitud máxima de 10, 9, 8, 7, 6, 5 ó 4 o menos nucleótidos.

Como se describe anteriormente y se ejemplifica más adelante en el Ejemplo III, se pueden identificar uno o más residuos de nucleótidos en el polinucleótido utilizando señales eléctricas obtenidas de dos etapas fraccionarias de un ciclo de translocación completo. El empleo de las señales de ambas etapas de translocación fraccional proporciona una señal duplicada para el mismo nucleótido o más nucleótidos y permite una mayor precisión dentro de una única determinación. Por consiguiente, emplear señales de ambas etapas de traducción fraccionada puede resultar en una mayor precisión de caracterización, con tasas de error reducidas entre un 25 y un 50% en comparación con la identificación de uno o más nucleótidos utilizando una única señal eléctrica u otra obtenida de un ciclo completo de translocación. De manera similar, el uso de señales de ambas etapas de conversión fraccional puede dar como resultado una reducción de las tasas de error en al menos 5%, 10%, 20%, 30%, 40%, 45%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% o superior. Dadas las enseñanzas y la guía que se proporcionan en el presente documento, los expertos en la materia sabrán cómo ajustar la precisión para un propósito determinado, como, por ejemplo, reduciendo el tamaño de la zona de constricción como se describe anteriormente para aumentar la resolución de la translocación de nucleótidos.

La información adicional obtenida de las etapas de translocación fraccional se puede usar para avanzar en la secuenciación de nanoporos de varias maneras. Por ejemplo, las mediciones obtenidas de las etapas de translocación fraccional para la misma palabra de nucleótido se pueden usar en algoritmos para mejorar la precisión de la resolución de la base de los nanoporos. Las mediciones obtenidas de las etapas de translocación fraccional para la misma palabra de nucleótido se pueden usar para reducir las tasas de error de lectura de homopolímeros porque la misma palabra de nucleótido se lee dos veces en una única determinación. En consecuencia, las mediciones obtenidas de las etapas de translocación fraccional para la misma palabra de nucleótido duplican la resolución resoluble de la respuesta de translocación de polinucleótido nativa, lo que da como resultado una resolución mejorada de patrones específicos de secuencia. Un uso de este último son los algoritmos de reconocimiento de patrones específicos de secuencia para detectar repeticiones de secuencia o polimorfismos de un solo nucleótido (SNP).

Como se estableció anteriormente, un método puede incluir (a) causar una diferencia de potencial a través de un poro en contacto con una helicasa Hel308 y un polinucleótido diana; (b) medir una o más señales producidas por una o más etapas de translocación fraccional del polinucleótido diana a través del poro, y (c) caracterizar el polinucleótido diana a partir de la señal eléctrica de las etapas de translocación fraccional. En algunos aspectos, el método incluye además repetir las etapas (a) - (c) una o más veces. Al repetir las etapas (a) - (c), se pueden caracterizar los nucleótidos adyacentes o las palabras de nucleótidos adyacentes. La repetición de las etapas (a) - (c) se puede realizar según se desee hasta que se caracterice parte o la totalidad del polinucleótido diana. Por ejemplo, la secuencia de una parte o la totalidad de un polinucleótido diana se puede determinar a través de cualquier número deseado de iteraciones de las etapas (a) - (c). Por consiguiente, se pueden determinar una o más características para la totalidad o parte de un polinucleótido diana.

Como se describe en el presente documento, cualquier helicasa Hel308 o variante de la misma se puede usar de acuerdo con las presentes realizaciones. Se presentan helicasas Hel308 ejemplares a continuación en las Tablas 1 y 2.

ES 2 735 015 T3

Tabla 1. Helicasas Hel308 ejemplares

Entrada	Descripción
NP_578406.1	helicasa tipo ski2 [Pyrococcus furiosus DSM 3638] >sp O73946.1 HEL5_PYRFU Nombre Rec: Completo=helicasa tipo ski2 supuesta >pdb 2ZJ2 A Cadena A, Estado Hjm Apo de ADN Helicasa en la Forma 1 >pdb 2ZJ5 A Cadena A, ADN Helicasa Hjm Arqueal Complejada con Adp en la Forma 1 >pdb 2ZJ8 A Cadena A, Estado Hjm Apo de ADN Helicasa en la Forma 2 >pdb 2ZJA A Cadena A, Helicasa Hjm de ADN Arqueal Complejada con Amppcp en la Forma 2 >dbj BAA32016.1 helicasa [Pyrococcus furiosus] >gb AAL80801.1 helicasa [Pyrococcus furiosus DSM 3638]
NP_126564.1	helicasa tipo ski2 [Pyrococcus abyssi GE5] >sp Q9V0A9.1 HEL5_PYRAB Nombre Rec: Completo=helicasa tipo ski2 supuesta >emb CAB49795.1 ADN helicasa [Pyrococcus abyssi GE5]
NP_143168.1	helicasa tipo ski2 [Pyrococcus horikoshii OT3] >sp O59025.1 HEL5_PYRHO Nombre Rec: Completo=helicasa tipo ski2 supuesta >dbj BAA30383.1 proteína hipotética de 715aa de longitud [Pyrococcus horikoshii OT3]
YP_004424773.1	helicasa tipo ski2 [Pyrococcus sp. NA2] >gb AEC52769.1 helicasa tipo ski2 [Pyrococcus sp. NA2]
YP_004623750.1	helicasa tipo ski2 [Pyrococcus yayanosii CHI] >gb AEH24478.1 helicasa tipo ski2 [Pyrococcus yayanosii CHI]
YP_002307730.1	helicasa tipo ski2 [Thermococcus onnurineus NA1] >gb ACJ16833.1 ADN helicasa [Thermococcus onnurineus NA1]
YP_004763427.1	helicasa tipo ski2 [Thermococcus sp. 4557] >gb AEK73750.1 helicasa tipo ski2 [Thermococcus sp. 4557]
YP_002959236.1	helicasa tipo ski2 [Thermococcus gammatolerans EJ3] >gb ACS33372.1 helicasa tipo ski2 supuesta [Thermococcus gammatolerans EJ3]
YP_004071709.1	helicasa tipo ski2 [Thermococcus barophilus MP] >gb ADT84486.1 helicasa tipo ski2 supuesta [Thermococcus barophilus MP]
YP_002994328.1	helicasa tipo ski2 supuesta [Thermococcus sibiricus MM 739] >gb ACS89979.1 helicasa tipo ski2 supuesta [Thermococcus sibiricus MM 739]
ZP_04875329.1	enzima de restricción Tipo III, familia subunidad res [Aciduliprofundum boonei T469] >gb EDY3511 1.1 enzima de restricción Tipo III, familia subunidad res [Aciduliprofundum boonei T469]
YP_003436565.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Ferroglobus placidus DSM 10642] >gb ADC66290.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Ferroglobus p1acidus DSM 10642]
YP_004485304.1	helicasa tipo ski2 [Methanoterris igneus Kol5] >gb AEF97239.1 helicasa tipo ski2 [Methanoterris igneus Kol 5]
YP_004616424.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanosalsum zhilinae DSM 4017] >gb AEH61205.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanosalsum zhilinae DSM 4017]
ZP_04873370.1	enzima de restricción Tipo III, familia subunidad res [Aciduliprofundum boonei T469] >ref YP_003482774.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Aciduliprofundum boonei T469] >gb EDY36687.1 enzima de restricción Tipo III, familia subunidad res [Aciduliprofundum boonei T469] >gb ADD08212.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Aciduliprofundum boonei T469]
YP_004342552.1	helicasa tipo ski2 [Archaeoglobus veneficus SNP6] >gb AEA47837.1 helicasa tipo ski2 [Archaeoglobus veneficus SNP6]
NP_071282.1	familia helicasa ski2 [Archaeoglobus fulgidus DSM 4304]

Entrada	Descripción
2P6R_A	Cadena A, Estructura del cristal de la Superfamilia 2 Helicasa Hel308 en Complejo con ADN desenrollado >pdb 2P6U A Cadena A, Estructura Apo de la Superfamilia Hel308 2 Helicasa
YP_685308.1	helicasa tipo ski2 [arquea metanogénica no cultivada RC-1] >sp Q0W6L1.1 HEL5_UNCMA Nombre Rec: Completo=helicasa tipo ski2 supuesta >emb CAJ35982.1 helicasa tipo ski2 supuesta [arquea metanogénica no cultivada RC-1]
YP_001048404.1	helicasa tipo ski2 [Methanoculleus marisnigri JR1 >gb ABN58422.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanoculleus marisnigri JR1]
YP_919908.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Thermophilum pendens Hrk 5] >gb ABL77905.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Thermophilum pendens Hrk 5]
YP_843229.1	helicasa tipo ski2 [Methanosaeta thermophila PT] >gb ABK14589.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanosaeta thermophila PT]
ZP_08045937.1	helicasa tipo ski2 [Haladaptatus paucihalophilus DX253] >gb EFW90585.1 helicasa tipo ski2 [Haladaptatus paucihalophilus DX253]
NP_280985.1	helicasa tipo ski2 [Halobacterium sp. NRC-1] >ref YP_001690117.1 helicasa tipo ski2 [Halobacterium salinarum R1] .sp Q9HMY6.1 HEL5_HALSA Nombre Rec: Completo=helicasa tipo ski2 supuesta>sp B0R7Q2.1 HEL5_HALS3 Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb AAG20465.1 proteína reparadora de ADN [Halobacterium sp. NRC-1] >emb CAP14771.1 supuesta ADN helicasa [Halobacterium salinarum R1]
YP_003357840.1	migración helicasa Holliday junction [Methanocella paludicola SANA] dbj BAI62857.1 migración helicasa Holliday junction [Methanocella paludicola SANA]
YP_003457479.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanocaldococcus sp. FS406-22] >gb ADC68743.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanocaldococcus sp. FS406-22]
YP_003127632.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanocaldococcus fervens AG86] >gb ACV24132.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanocaldococcus fervens AG86]
YP_003735335.1	helicasa tipo ski2 [Halalkalicoccus jeotgali B3] >gb ADJ13543.1 helicasa tipo ski2 [Halalkalicoccus j eotgali B3]
YP_503885.1	helicasa tipo ski2 [Methanospirillum hungatei JF-1] >gb ABD42166.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína tipo helicasa [Methanospirillum hungatei JF-1]
BAJ48115.1	helicasa [Candidatus Caldiarchaeum subterraneum] >dbj BAJ48144.1 helicasa [Candidatus Caldiarchaeum subterraneum] >dbj BAJ50919.1 helicasa [Candidatus Caldiarchaeum subterraneum]
YP_001405615.1	helicasa tipo ski2 [Candidatus Methanoregula boonei 6A8] >sp A71B61.1 HEL5_METB6 Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb ABS56972.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanoregula boonei 6A8]
YP_306959.1	helicasa tipo ski2 [Methanosarcina barkeri str. Fusaro] >sp Q465R3.1 HEL5_METBF Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb AAZ72379.1 helicasa [Methanosarcina barkeri str. Fusaro]
YP_001031179.1	helicasa tipo ski2 [Methanocorpusculum labreanum Z] >gb ABN07912.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanocorpusculum labreanum Z]
YP_003541733.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Methanohalophilus mahii DSM 5219] >gb ADE36088.1 DEAD(SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanohalophilus mahii DSM 5219]

Entrada	Descripción
YP_004384692.1	supuesta helicasa Ski2 [Methanosaeta concilii GP6] >gb AEB68874.1 supuesta helicasa Ski2 [Methanosaeta concilii GP6]
YP_003725904.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanohalobium evestigatum Z-7303] >gb AD173108.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanohalobium evestigatum Z-7303]
YP_003405271.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Haloterrigena turkmenica DSM 5511] >gb ADB62598.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Haloterrigena turkmenica DSM 5511]
YP_004244914.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Vulcanisaeta moutnovskia 768-28] >gb ADY01412.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Vulcanisaeta moutnovskia 768-28]
YP_001540156.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Caldivirga maquilingensis IC-167] >sp A8MB76.1 HELS_CALMQ Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb ABW01166.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Caldivirga maquilingensis IC-167]
NP_618094.1	helicasa tipo ski2 [Methanosarcina acetivorans C2A] >sp Q8TL39.1 HELS_METAC Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb AAM06574.1 helicasa [Methanosarcina acetivorans C2A]
YP_003900980.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Vulcanisaeta distributa DSM 14429] >gb ADN49929.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Vulcanisaeta distributa DSM 14429]
YP_003896003.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanoplanus petrolearius DSM 11571] >gb ADN37565.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanoplanus petrolearius DSM 11571]
YP_003615773.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [methanocaldococcus infernus ME] >gb ADG12809.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanocaldococcus infernus ME]
YP_183745.1	proteína RNA helicasa tipo Ski2 [Thermococcus kodakarensis KOD1] >sp Q5JGV6.1 HELS_PYRKO Nombre Rec: Completo=helicasa tipo ski2 supuesta: Contiene: Nombre Rec: Completo=Endonucleasa P1-PkoHel; Nombre Alt: Completo=Pko Hel intein >dbj BAD85521.1 homólogo de RNA helicasa Ski2 [Thermococcus kodakarensis KOD1]
YP_001322557.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus vannielii SB] >sp A6UN73.1 HELS_METVS Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb ABR53945.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanococcus vannielii SB]
YP_002467772.1	helicasa tipo ski2 [Methanosphaerula palustris E1-9c] >gb ACL18049.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanosphaerula palustris E1-9c]
YP_003480097.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Natrialba magadii ATCC 43099] >gb ADD05535.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Natrialba magadii ATCC 43099]
YP_004577043.1	helicasa tipo ski2 [Methanothermococcus okinawensis IH1] >gb AEH07265.1 helicasa tipo ski2 [Methanothermococcus okinawensis IH1]
YP_004742641.1	helicasa superfamilia II [Methanococcus maripaludis XI] >gb AEK19898.1 helicasa superfamilia II [Methanococcus maripaludis XI]
NP_632449.1	helicasa tipo ski2 [Methanosarcina mazei Go1] >sp Q8PZR7.1 HELS_METMA Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb AAM30121.1 helicasa [Methanosarcina mazei Go1]

Entrada	Descripción
YP_001097223.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus maripaludis C5 >gb ABO35008.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanococcus maripaludis C5]
YP_004742247.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus maripaludis XI] >gb AEK19504.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus maripaludis XI]
YP_004794766.1	helicasa tipo ski2 [Haloarcula hispanica ATCC 33960] >gb AEM55778.1 helicasa tipo ski2 [Haloarcula hispanica ATCC 33960]
NP_988010.1	helicasa superfamilia II [Methanococcus maripaludis S2] >emb CAF30446.1 helicasa superfamilia II [Methanococcus maripaludis S2]
YP_565780.1	helicasa tipo ski2 [Methanococcoides burtonii DSM 6242] >sp Q12WZ6.1 HELS_METBU Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb ABE52030.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína tipo helicasa [Methanococcoides burtonii DSM 6242]
YP_001549808.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus maripaludis C6] >gb ABX02576.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanococcus maripaludis C6]
YP_001548609.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus maripaludis C6] >gb ABX01377.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanococcus maripaludis C6]
YP_001329359.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus maripaludis C7] >gb ABR65208.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanococcus maripaludis C7]
YP_004595982.1	helicasa tipo ski2 [Halopiger xanaduensis SH-6] >gb AEH36103.1 helicasa tipo ski2 [Halopiger xanaduensis SH-6]
YP_656795.1	helicasa tipo ski2 [Haloquadratum walsbyi DSM 16790] >emb CAJ51138.1 ADN helicasa dependiente de ATP [Haloquadratum walsbyi DSM 16790]
CCC38992.1	ADN helicasa dependiente de ATP Hel308 [Haloquadratum walsbyi C23]
YP_004035272.1	helicasa superfamilia II [Halogeometricum borinquense DSM 11551]
YP_137330.1	helicasa tipo ski2 [Haloarcula marismortui ATCC 43049] >sp Q5UYM9, 1 HELS_HALMA Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb AAV47624.1 helicasa tipo ski2 supuesta [Haloarcula marismortui ATCC 43049]
YP_001581577.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Nitrosopumilus maritimus SCM] >gb ABX12139.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Nitrosopumilus maritimus SCM1]
EET90255.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Candidatus Micrarchaeum acidiphilum ARMAN-2]
NP_376477.1	Helicasa [Sulfolobus tokodaii str. 7 >sp Q974S1.1 HELS_SULTO Nombre Rec: Completo=helicasa tipo ski2 supuesta>dbj BAK54341.1 migración helicasa Holliday junction [Sulfolobus tokodaii str. 7]
YP_001097792.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene dominio helicasa [Methanococcus maripaludis C5] >gb ABO35578.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanococcus maripaludis C5]
ZP_08667240.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Nitrosopumilus sp. MYI] >gb EGP92972.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Nitrosopumilus sp. MYI]

Entrada	Descripción
YP_254972.1	ADN helicasa [Sulfolobus acidocaldarius DMS 639] >sp Q4JC00.1 HELS_SULAC Nombre Rec: Completo=helicasa tipo ski2 supuesta >gb AAY79679.1 ADN helicasa [Sulfolobus acidocaldarius DSM 6391]
EFD92533.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Candidatus Parvarchaeum acidophilus ARMAN-5]
YP_003176257.1	helicasa tipo ski2 [Halomicrobium mukohataei DSM 12286] >gb ACV46820.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halomicrobium mukohataei DSM 12286]
EGD71904.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína dominio helicasa [Candidatus Parvarchaeum acidophilus ARMAN-5_'5-way FS']
YP_001040230.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Staphylothermus marinus F1] >gb ABN69322.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Staphylothermus marinus F1]
ABZ07376.1	supuesta DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [no cultivada marine crenarchaeote HF400_AN1W133M9]
YP_001097458.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanoccus maripaludis C5] >gb AB035243.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanoccus maripaludis C5]
ABZ08606.1	supuesta DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [no cultivada marine crenarchaeote HF4000_APKG3H9]
YP_325906.1	helicasa tipo ski2 [Natronomonas pharaonis DSM 2160] >sp Q31U46.1 HELS_NATPD Nombre Rec: Completo=helicasa tipo ski2 supuesta >emb CA148337.1 ADN helicasa dependiente de ATP [Natronomonas pharaonis DSM 2160]
YP_930665.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Pyrobaculum islandicum DSM 4184] >gb ABL88322.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Pyrobaculum islandicum DSM 4184]
YP_001435870.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Ignicoccus hospitalis KIN4/1] >gb ABU82463.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Ignicoccus hospitalis KIN4/1]
YP_003668634.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Staphylothermus hellenicus DSM 12710] >gb AD131735. DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Staphylothermus hellenicus DSM 12710]
ZP_08558598.1	helicasa tipo ski2 [Halorhabdus tiamatea SARL4B] >g EGM36528.1 helicasa tipo ski2 [Halorhabdus tiamatea SARL4B]
YP_002428409.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Desulfurococcus kamchatkensis 1221n] >gb ACL11042.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Desulfurococcus kamchatkensis 1221n]
YP_004336918.1	helicasa de unión de ADN dependiente de ATP [Thermoproteus uzoniensis 768-20] >gb AEA11606.1 helicasa de unión de ADN dependiente de ATP [Thermoproteus uzoniensis 768-20]
ZP_08257442.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Candidatus Nitrosoarchaeum limnia SFB1] >gb EGG41989.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Candidatus Nitrosoarchaeum limnia SFB1]
YP_004459284.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Acidianus hospitalis W1] >gb AEE94986.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Acidianus hospitalis W1]

Entrada	Descripción
NP_558924.1	helicasa de unión de ADN dependiente de ATP [Pyrobaculum aerophilum str. IM2] > gb AAL63106.1 helicasa de unión de ADN dependiente de ATP [Pyrobaculum aerophilum str. IM2]
YP_004409449.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Metallosphaera cuprina Ar-4] > gb AEB94965.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Metallosphaera cuprina Ar-4]
YP_003649556.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Thermosphaera aggregans DSM 11486] > gb ADG90604.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Thermosphaera aggregans DSM 11486]
ZP_06387115.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus solfataricus 98/2] > gb ACX90562.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus solfataricus 98/2]
2VA8_A	Cadena A, helicasa de reparación de ADN Hel308>pdb 2VA8 B Cadena B, helicasa de reparación de ADN Hel308>emb CAO85626.1 ADN helicasa [Sulfolobus solfataricus]
YP_004809267.1	helicasa tipo ski2 [halophilic arquea DL31]>gb AEN06894.1 helicasa tipo ski2 [halophilic arquea DL31]
ADX84345.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus REY15A]>gb ADX81629.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus HVE10/4]
YP_002828439.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Sulfolobus islandicus M.14.25] >ref YP_002842325.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus M.16.27]>gb ACP37141.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus M.14.25]>gb ACP54280.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus M.16.27]
YP_002913571.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus M.16.4]>gb ACR40903.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus M.16.4]
Q97VY9.1	Nombre Rec: Completo=helicasa tipo ski2 supuesta
YP_002841682.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus Y.N.15.51]>gb ACP49760.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus Y.N.1 5.51]
YP_002831080.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus L.S.2.15]>ref YP_003418425.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus L.D.8.5]>gb ACP34435.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus L.S.2.15]>gb ADB86055.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus L.D.8.5]
YP_001054984.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Pyrobaculum calidifontis JCM 11548]>sp A3MSA.1.1 HELs_PYRCJ Nombre Rec: Completo=helicasa tipo ski2 supuesta>gb ABO07518.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Pyrobaculum calidifontis JCM 11548]
NP_343811.1	proteína relacionada con ADN helicasa [Sulfolobus solfataricus P2]>ref YP_002836469.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Sulfolobus islandicus Y.G.57.14]>gb AAK42601.1 proteína relacionada con ADN helicasa [Sulfolobus solfataricus P2]>gb ACP44547.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sulfolobus islandicus Y.G.57.14]
YP_001152379.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Pyrobaculum arsenaticum DSM 13514] > gb ABP49727.1 DEAD (SEC ID NO: 2)/DEAH (SEC

Entrada	Descripción
	ID NO: 3) cuadro proteína de dominio de helicasa [Pyrobaculum arsenaticum DSM 13514]
YP_001191456.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Metallosphaera sedula DSM 5348] > gb ABP95532.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Metallosphaera sedula DSM 5348]
NP_147034.2	migración helicasa holliday junction [Aeropyrum pernix K1] > sp Q9YFQ8.2 HELIS_AERPE Nombre Rec: Completo=helicasa tipo ski2 supuesta > dbj BAA79103.2 migración helicasa holliday junction [Aeropyrum pernix K1]
YP_024158.1	helicasa tipo ski2 [Picrophilus torridus DSM 9790] > gb AAT43965.1 helicasa implicada en la protección UV [Picrophilus torridus DSM 9790]
YP_003816358.1	helicasa tipo ski2 supuesta [Acidilobus saccharovorans 345-15] > gb ADL19327.1 helicasa tipo ski2 supuesta [Acidilobus saccharovorans 345-15]
YP_003860265.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Ignisphaera aggregans DSM 17230] > gb ADM28385.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Ignisphaera aggregans DSM 17230]
NP_394295.1	helicasa tipo ski2 [Thermoplasma acidophilum DSM 1728] > sp Q9HJX7.1 HELIS_THEAC Nombre Rec: Completo=helicasa tipo ski2 supuesta > emb CAC11964.1 proteína relacionada con ADN helicasa [Thermoplasma acidophilum]
YP_876638.1	helicasa superfamilia II [Cenarchaeum symbiosum A] > gb ABK78334.1 helicasa superfamilia II [Cenarchaeum symbiosum A]
ZP_05571398.1	helicasa tipo ski2 [Ferroplasma acidarmanus fer1]
YP_004176252.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Desulfurococcus mucosus DSM 2162] > gb ADV64770.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Desulfurococcus mucosus DSM 2162]
YP_001737782.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Candidatus Korarchaeum cryptofilum OPF8] > gb ACB08099.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Candidatus Korarchaeum cryptofilum OPF8]
EGQ40435.1	helicasa superfamilia II [Candidatus Nanosalinarum sp. J07AB56]
YP_002567343.1	helicasa tipo ski2 [Halorubrum lacusprofundi ATCC 49239] > gb ACM58273.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorubrum lacusprofundi ATCC 49239]
YP_001793507.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Thermoproteus neutrophilus V24Sta] > gb ACB39061.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Thermoproteus neutrophilus V24Sta]
YP_003534088.1	ADN helicasa dependiente de ATP Hel308a [Haloferax volcanii DS2] > gb ADE04048.1 ADN helicasa dependiente de ATP Hel308a [Haloferax volcanii DS2]
YP_004037165.1	helicasa Superfamilia II [Halogeometricum borinquense DSM 11551] > gb ADQ67720.1 helicasa superfamilia II [Halogeometricum borinquense DSM 11551]
NP_111333.1	helicasa tipo ski2 [Thermoplasma volcanium GSS1] > sp Q97AI2.1 HELIS_THEVO Nombre Rec: Completo=helicasa tipo ski2 supuesta > dbj BAB59970.1 ADN helicasa [Thermoplasma volcanium GSS1]
YP_002565871.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Halorubrum lacusprofundi ATCC 49239] > gb ACM56801.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorubrum lacusprofundi ATCC 49239]
CCC39675.1	ADN helicasa dependiente de ATP Hel308 [Haloquadratum walsbyi C23]

Entrada	Descripción
YP_657401.1	ADN helicasa dependiente de ATP [Haloquadratum walsbyi DSM 16790] >emb[CAJ51759.1] ADN helicasa dependiente de ATP [Haloquadratum walsbyi DSM 16790]
YP_003535028.1	ADN helicasa dependiente de ATP Hel308b [Haloferax volcanii DS2]>gb ADE02398.1 ADN helicasa dependiente de ATP HEL308b [Haloferax volcanii DS2]
YP_003706863.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus voltae A3]>gb ADI35890.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanococcus voltae A3]
ABD 17736.1	helicasa [Methanococcus voltae PS]
NP_613398.1	helicasa superfamilia II [Methanopyrus kandleri AV19]>gb AAM01328.1 helicasa Superfamilia II predicha [Methanopyrus kandleri AV19]
CBH38575.1	helicasa tipo ski2 supuesta [arquea no cultivada]
EEZ93258.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Candidatus Parvarchaeum acidiphilum ARMAN-4]
YP_004004246.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanothermus fervidus DSM 2088]>gb ADP77484.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanothermus fervidus DSM 2088]
YP_003850109.1	helicasa [Methanothermobacter marburgensis str. Marburg]>gb ADL58796.1 helicasa predicha [Methanothermobacter marburgensis str. Marburg]
YP_003424423.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanobrevibacter ruminantium M1]>gb ADC47531.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanobrevibacter ruminantium M1]
YP_004291107.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanobacterium sp. AL-21]>gb ADZ10135.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanobacterium sp AL-21]
YP_447162.1	helicasa [Methanosphaera stadtmanae DSM 3091] >gb ABC56519.1 helicasa predicha [Methanosphaera stadtmanae DSM 3091]
YP_004519549.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanobacterium sp. SWAN-1]>gb AEG17748.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanobacterium sp. SWAN-1]
NP_275949.1	proteína relacionada con ADN helicasa [Methanothermobacter thermotrophicus str. Delta H]>sp O26901.1 HELS...METTH Nombre Rec: Completo=helicasa tipo ski2 supuesta>gb AAB85310.1 proteína relacionada con ADN helicasa [Methanothermobacter thermotrophicus str. Delta H]
ZP_05975717.2	supuesta helicasa Ski2 [Methanobrevibacter smithii DSM 2374]>gb EFC93382.1 supuesta helicasa Ski2 [Methanobrevibacter smithii DSM 2374]
ZP_03607647.1	proteína hipotética METSMIALI...00751 [Methanobrevibacter smithii DSM 2375]>gb EEE41862.1 proteína hipotética METSMIALI_00751 [Methanobrevibacter smithii DSM 2375]
YP_001273412.1	helicasa dependiente de ATP [Methanobrevibacter smithii ATCC 35061]>gb ABQ87044.1 helicasa dependiente de ATP [Methanobrevibacter smithii ATCC 35061]
YP_003247505.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanocaldococcus vulcanius M7] >gb ACX73023.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanocaldococcus vulcanius M7]
NP_248116.1	familia helicasa tipo ski2 [Methanocaldococcus jannaschii DSM 2661] >sp Q58524.1 HELS...METJA Nombre Rec: Completo=helicasa tipo ski2 supuesta; Contiene: Nombre Rec: Completo=Endonucleasa PI-MjaHel; Nombre Alt: Completo=Mja Hel intein; Nombre Alt:

Entrada	Descripción
	Completo=Mja Pep3 intein>gb AAB99126.1 supuesta familia helicasa ski2 [Methanocaldococcus jannaschii DSM 26621]
YP_001324295.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Methanococcus acolicus Nakai-3]>gb ABR 55683.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Methanococcus aeolicus Nankai-3]
YP_003536960.1	helicasa de división pre-mRNA [Haloferax volcanii DS2]>gb ADE02332.1 helicasa de división pre-mRNA [Haloferax volcanii DS2]
YP_003131029.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorhabdus utahensis DSM 12940]>gb ACV12296.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorhabdus utahensis DSM 12940]
YP_002567151.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Halorubrum lacusprofundi ATCC 49239] >gb ACM58081.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorubrum lacusprofundi ATCC 49239]
YP_004035351.1	helicasa superfamilia II [Halogeometricum borinquense DSM 1151] >gb ADQ65912.1 helicasa superfamilia II [Halogeometricum borinquense DSM 11551]
YP_004808851.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [halophilic arquea DL31]>gb AEN06478.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [halophilic arquea DL31]
XP_002716686.1	predicha:ADN polimerasa isoforma 1 theta [Oryctolagus cuniculus]
YP_656834.1	ADN helicasa dependiente de ATP [Haloquadratum walsbyi DSM 16790] >emb CAJ51176.1 ADN helicasa dependiente de ATP [Haloquadratum walsbyi DSM 16790]
XP_003248103.1	predicha:ADN polimerasa isoforma 1 tipo theta [Acyrtosiphon pisum]
ABC72356.1	ADN helicasa dependiente de ATP [Haloquadratum walsbyi]
CCC39031.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Haloquadratum walsbyi C23]
XP_001165150.2	predicha:ADN polimerasa isoforma 1 theta [Pan troglodytes]
XP_003225852.1	predicha:ADN polimerasa tipo theta [Anolis carolinensis]
XP_615375.3	predicha:ADN polimerasa theta [Bos Taurus]>ref[XP_002684835.1] predicha:polimerasa (ADN-dirigida), tipo theta [Bos Taurus] >gb DAA33456.1 polimerasa (ADN-dirigida), tipo theta [Bos Taurus]
XP_002813286.1	predicha:proteína de baja calidad:ADN polimerasa tipo theta [Pongo abelii]
AAR08421.2	ADN polimerasa theta [Homo sapiens]
EAW79510.1	polimerasa (ADN-dirigida) theta, isoforma CRA a [Homo sapiens]
NP_955452.3	ADN polimerasa theta [Homo sapiens]>sp1075417.21DPOLQ...HUMAN Nombre Rec: Completo=ADN polimerasa theta; Nombre Alt: Completo=theta ADN polimerasa >gb aa172289.1 polimerasa (ADN-dirigida), theta [polinucleótido sintético]
NP_001099348.1	ADN polimerasa theta [Rattus norvegicus]>gb EDM11249.41 polimerasa (ADN-dirigida), theta (predicha), isoforma CRA a [Rattus norvegicus]
XP_003341262.1	predicha:proteína de baja calidad:ADN polimerasa tipo theta [Monodelphis domestica]
XP_001502374.3	predicha:ADN polimerasa theta [Equus caballus]
XP_545125.3	predicha:proteína de baja calidad:ADN polimerasa theta [Canis lupus familiaris]

Entrada	Descripción
XP_002928855.1	predicha: proteína de baja calidad: ADN polimerasa tipo theta [Ailuropoda melanoleuca]
NP_084253.1	ADN polimerasa theta isoforma 1 [MUS musculus]>gb AAL77225.1 ADN polimerasa theta [Mus musculus]>gb EDK_97951.1 polimerasa (ADN-dirigida), theta, isoforma CRA_a [Mus musculus]>gb AA138361.1 polimerasa (ADN-dirigida), theta [Mus musculus]>gb AA157901.1 polimerasa (ADN-dirigida), theta [Mus musculus]
AAK39635.1	ADN polimerasa theta [Homo sapiens]
AAN393838.1	ADN polimerasa Q [Mus musculus]
XP_003412882.1	predicha: ADN polimerasa theta [Loxodona africana]
YP_003735206.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Halalkalicoccus jeotgali B3]>gb ADJ13414.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halalkalicoccus jeotgali B3]
YP_004794841.1	helicasa de división pre-mRNA [Haloarcula hispanica ATCC 33960] >gb AEM55853.1 helicasa de división pre-mRNA [Haloarcula hispanica ATCC 33960]
XP_416549.2	predicha: ADN polimerasa tipo theta [Gallus gallus]
XP_003427319.1	PREDICTED: isoforma 2 tipo helicasa POLQ [Nasonia vitripennis]
XP_003202748.1	predicha:ADN polimerasa tipo theta [Meleagris gallopavo]
XP_969311.1	predicha: ADN polimerasa tipo theta [Tribolium castaneum]>gb EEZ97532.1 proteína hipotética TcasGA2..TC011380 [Tribolium castaneum]
ZP_08046037.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Haladaptatus paucihalophilus DX253]>gb EFW90685.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Haladaptatus paucihalophilus DX253]
YP_461714.1	helicasa [Syntrophus aciditrophicus SB]>gb ABC77546.1 helicasa [Syntrophus aciditrophicus SB]
YP_003176510.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Halomicrobium mukohataei DSM 12286]>gb AAV47694.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halomicrobium mukohataei DSM 12286]
YP_137400.1	helicasa de división pre-mRNA [Haloarcula marismortui ATCC43049]>gb AAV47694.1 helicasa de división pre-mRNA [Haloarcula marismortui ATCC 43049]
NP_001184156.1	polimerasa (ADN-dirigida), theta [Xenopus (Silurana) tropicalis]
NP_280861.1	helicasa de división pre-mRNA [Halobacterium sp. NRC-1]>ref YP...001689987.1 ADN helicasa dependiente de ATP [Halobacterium salinarum R1]>gb AAG20341.1 helicasa de división pre-mRNA [Halobacterium sp. NRC-1]>emb CAP14641.1 ADN helicasa dependiente de ATP [Halobacterium salinarum R1]
YP_004595640.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Halopiger xanaduensis SH 6]>gb AEH35761.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halopiger xanadueniss SH-6]
XP_001521144.2	predicha:ADN polimerasa theta, partial [Ornithorhynchus anatinus]
XP_003261953.1	predicha: ADN polimerasa theta, partial [Nomascus leucogenys]
XP_001358456.2	GA 19301 [Drosophila pseudoobscura pseudoobscura]>gb EAI...27595.2 GA 19301 [Drosophila pseudoobscura pseudoobscura]
ZP_0856003.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorhabdus tiamatea SARL4B]>gb EGM34502.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorhabdus tiamatea SARL4B]

Entrada	Descripción
XP_002187783.1	predicha: similar a polimerasa (ADN-dirigida), theta [Taeniopygia guttata]
XP_002112587.1	proteína hipotética TRIADDRAFT ...25163 [Trichoplax adhaerens]>gb EDV24697.1 proteína hipotética TRIADDRAFT...25163 [Trichoplax adhaerens]
YP_003405139.1	DEATH/DEAH cuadro helicasa [Haloterrigena turkmenica DSM 5511] >gb ADB62466.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Haloterrigena turkmenica DSM 5511]
EGV92665.1	ADN polimerasa theta [Cricetulus griseus]
CBY24305.1	producto proteico no nombrado [Oikopleura dioica]
YP_003130565.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorhabdus utahensis DSM 12940]>gb ACV11832.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Halorhabdus utahensis DSM 12940]
YP_003479811.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Natrialba magadii ATCC 43099]>gb ADD05249.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Natrialba magadii ATCC 43099]
EFB22383.1	proteína hipotética PANDA_000253 [Ailuropoda melanoleuca]
YP_003357334.1	supuesta helicasa dependiente de ATP [Methanocella paludicola SANA E]>dbj BA162351.1 supuesta helicasa dependiente de ATP [Methanocella paludicola SANA E]
YP_325942.1	ADN helicasa dependiente de ATP 2 [Natronomonas pharaonis DSM 2160]>emb CA148373.2 ADN helicasa dependiente de ATP 2 [Natronomonas pharaonis DSM 2160]
XP_002912509.1	predicha: proteína de baja calidad: helicasa tipo POLQ [Ailuropoda melanoleuca]
XP_002704678.1	predicha: helicasa, tipo POLQ [BOS Taurus]
CAE47762.2	nueva proteína similar a polimerasa theta humana dirigida por ADN (POLQ) [DANIO rerio]
XP_003205636.1	predicha: helicasa tipo POLQ [Meleagris gallopavo]
XP_544959.2	predicha: helicasa, tipo POLQ [Canis lupus familiaris]
EFX86757.1	proteína hipotética DAPPUDRAFT 312857 [Daphnia pulex]
YP_003389641.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro helicasa [Spirosoma lingual DSM 74]>gb ADB40842.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Spirosoma lingual DSM 74]
XP_002602932.1	proteína hipotética BRAFLDRAFT_251779 [Branchiostoma floridae]>gb EEN58944.1 proteína hipotética BRAFLDRAFT_251779 [Branchiostoma floridae]
YP_004144962.1	subunidad p20 catalítica de peptidasa C14 caspasa [Mesorhizobium ciceri biovar biserrulae WSM1271]>ref YP 004614892.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína que contiene el dominio de helicasa [Mesorhizobium opportunistum WSM2075] >gb ADV 14912.1 subunidad p20 catalítica de peptidasa C14 caspasa [Mesorhizobium ciceri biovar biserrulae WSM1271] >gb AEH90798.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Mesorhizobium opportunistum WSM2075]
XP_002124758.1	predicha: similar a ADN polimerasa theta [Ciona intestinalis]
XP_694437.5	predicha: ADN polimerasa theta [Danio rerio]
XP_420565.1	predicha: similar a ADN helicasa HEL308 [Gallus gallus]
XP_003129397.1	predicha: helicasa tipo POLQ [Sus scrofa]

Entrada	Descripción
EDL20278.1	mCG128467, isoforma CRA_b [Mus musculus]
XP_001517710.2	predicha: helicasa POLQ, parcial [Ornithorhynchus anatinus]
AAH82601.1	Helicasa, tipo mus 308 (Drosophila) [Mus musculus]
XP_003384429.1	predicha:ADN polimerasa tipo theta [Amphimedon queenslandica]
XP_003221282.1	predicha: helicasa tipo POLQ [Anolis carolinensis]
NP_524333.1	308 sensible a mutágeno [Drosophila melanogaster] >gb AAB67306.1 Mus308 [Drosophila melanogaster] >gb ACH92234.1 FI03732p [Drosophila melanogaster]
AAX33507.1	LP14642p [Drosophila melanogaster]
NP_001074576.1	helicasa tipo POLQ [Mus musculus >sp Q2VPA6.2 HELQ_MOUSE Nombre Rec: Completo=Helicasa tipo POLQ; Nombre Alt: Completo=helicasa tipo Mus 308: Nombre Alt: helicasa POLQ tipo completa>gb AAI09171.2 Helicasa, tipo mus 308 (Drosophila) [Mus musculus]
YP_003523727.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sideroxydans lithotrophicus ES-1] >gb ADE11340.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sideroxydans litotrophicus ES-1]
XP_002120889.1	predicha: similar a ADN helicasa HEL308 [Ciona intestinalis]
XP_001892566.1	enzima de restricción Tipo III, familia subunidad proteína res [Brugia malayi] >gb EDP38603.1 enzima de restricción Tipo III, familia subunidad proteína res [Brugia malayi]
ABZ09232.1	proteína del dominio C-terminal conservado de la supuesta helicasa [no cultivada marine crenarchaeote HF4000_APKG7F11]
XP_002814981.1	predicha: proteína de baja calidad: helicasa tipo POLQ [Pongo abelii]
XP_002717082.1	predicha: ADN helicasa HEL308 [Oryctolagus cuniculus]
XP_001104832.1	predicha: helicasa, tipo POLQ [Macaca mulatta]
AAL85274.1	ADN helicasa HEL308 [Homo sapiens]
NP_598375.2	helicasa tipo POLQ [Homo sapiens] >gb EAX05934.1 ADN helicasa HEL308, isoforma CRA_a[Homo sapiens] >gb AA141525.1 Helicasa, tipo POLQ [polinucleótido sintético]
Q8TDG4.2	Nombre Rec: Completo=Helicasa tipo POLQ; Nombre Alt: Completo=helicasa tipo Mus 308; Nombre Alt: Completo=tipo POLQ helicasa
XP_003265889.1	predicha: helicasa POLQ [Nomascus leucogenys]
XP_002745688.1	predicha: helicasa tipo POLQ [Callithrix jacchus]
XP_003310356.1	predicha: proteína de baja calidad: helicasa tipo POLQ [Pan troglodytes]
NP_001014156.2	helicasa, tipo POLQ [Rattus norvegicus] >ref XP_001060858.1 predicha: helicasa, tipo POLQ [Rattus norvegicus] >gb EDL99554.1 RCG37823, isoforma CRA_c [Rattus norvegicus]
XP_001850567.1	ADN helicasa dependiente de ATP MER3 [Culex quinquefasciatus] >gb EDS32308.1 ADN helicasa dependiente de ATP MER3 [Culex quinquefasciatus]
XP_003427318.1	predicha: helicasa tipo POLQ isoforma 1 [Nasonia vitripennis]
XP_003143912.1	proteína hipotética LOAG_08332 [Loa loa] >gb EFO20157.1 proteína hipotética LOAG_08332 [Loa loa]
CAG11187.1	producto proteico no nombrado [Tetraodon nigroviridis]

ES 2 735 015 T3

Entrada	Descripción
XP_001111254.2	predicha:ADN polimerasa theta isoforma 2 [Macaca mulatta]
XP_003414242.1	predicha: helicasa POLQ [Loxodonta africana]
XP_002681870.1	proteína predicha [Naegleria gruberi] >gb EFC49126.1 proteína predicha [Naegleria gruberi]
EAX05935.1	ADN helicasa HEL308, isoforma CRA_b [Homo sapiens]
AAH59917.1	proteína Ascc3 [Mus musculus]
ZP_07082808.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sphingobacterium spiritivorum ATCC 33861] >gb EFK55937.1 DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Sphingobacterium spiritivorum ATCC 33861]
XP_001494572.3	predicha: proteína de baja calidad: helicasa tipo POLQ [Equus caballus]
XP_002714920.1	predicha: subunidad 3 del complejo cointegrador 1 activador de señal [Oryctolagus cuniculus]
XP_002598278.1	proteína hipotética BRAFLDRAFT_204526 [Branchiostoma floridae] >gb EEN54290.1 proteína hipotética BRAFLDRAFT_204526 [Branchiostoma floridae]
XP_001943294.1	predicha: helicasa tipo POLQ isoforma 1 [Acyrtosiphon pisum] >ref XP_003240510.11 predicha: isoforma 2 tipo helicasa POLQ [Acyrtosiphon pisum]
XP_002803889.1	predicha: tipo subunidad 3 del complejo cointegrador 1 activador de señal [Macaca mulatta]
XP_001651546.1	ADN polimerasa theta [Aedes aegypti] >gb EAT42599.1 ADN polimerasa theta [Aedes aegypti]
CAA11679.1	ARN helicasa [Homo sapiens]
XP_002837795.1	proteína hipotética [Tuber melanosporum Mel28] >emb CAZ1986.1 producto proteico no nombrado [Tuber melanosporum]
EGT47882.1	proteína hipotética CAEBREN_02542 [Caenorhabditis brenneri]
EDL99655.1	subunidad 3 del complejo cointegrador 1 activador de señal (predicha), isoforma CRA b [Rattus norvegicus]
NP_932124.2	subunidad 3 del complejo cointegrador 1 activador de señal [Mus musculus]
EDL05054.1	mCG1 19534 [Mus musculus]
gij352115865 ZP_08963952.1	DEAD (SEC ID NO: 2)/DEAH (SEC ID NO: 3) cuadro proteína de dominio de helicasa [Natriema pellirubrum DSM 15624]

En la Tabla 2 a continuación se muestran más realizaciones de las helicasas Hel308, así como los motivos Hel308 y los motivos Hel308 extendidos.

Tabla 2. Helicasas Hel308 ejemplares, motivos Hel308 y motivos Hel308 ampliados.

NO. GI	Helicasa	Nombres	% Identidad	% Identidad	Motivo Hel308	Motivo Hel308 ampliado
			Hel308 Pfu	Hel308 Mbu		
GI:121689265	Hel308 Mbu	Methanococcoides burtonii	37%		QMAGRAGR (SEC ID NO: 4)	QMAGRAGRP (SEC ID NO: 5)
GI:18202135	Hel308 Pfu	Pyrococcus furiosus DSM 3638		37%	QMLGRAGR (SEC ID NO: 6)	QMLGRAGRP (SEC ID NO: 7)
GI:490144641	Hel308 Hvo	Haloferax volcanii	34%	41%	QMMGRAGR (SEC ID NO: 8)	QMMGRAGRP (SEC ID NO: 9)
GI:506391664	Hel308 Hla	Halorubrum lacusprofundi	35%	42%	QMCGRAGR (SEC ID NO: 10)	QMGRAGRP (SEC ID NO: 11)
GI:118576895	Hel308 Csy	Cenarchaeum symbiosum	34%	34%	QLCGRAGR (SEC ID NO: 12)	QLCGRAGRP (SEC ID NO: 13)
GI:18202627	Hel308 Sso	Sulfolobus solfataricus	35%	33%	QMSGGRAGR (SEC ID NO: 14)	QMSGGRAGRP (SEC ID NO: 15)
**	Hel308 Mfr	Methanogenium frigidum	37%	44%	QMAGRAGR (SEC ID NO: 16)	QMAGRAGRP (SEC ID NO: 17)
GI:503633371	Hel308 Mok	Methanothermococcus okinawensis	37%	34%	QCIGRAGR (SEC ID NO: 18)	QCIGRAGRP (SEC ID NO: 19)
GI:333911571	Hel308 Mig	Methanoterris igneus Kol 5	40%	35%	QCIGRAGR (SEC ID NO: 20)	QCIGRAGRP (SEC ID NO: 21)
GI:240102927	Hel308 Tga	Thermococcus gammatolerans EJ3	60%	38%	QMMGRAGR (SEC ID NO: 22)	QMMGRAGRP (SEC ID NO: 23)
GI:315231273	Hel308 Tba	Thermococcus barophilus MP	57%	35%	QMIGRAGR (SEC ID NO: 24)	QMIGRAGRP (SEC ID NO: 25)
GI:242398904	Hel308 Tsi	Thermococcus sibiricus MM 739	56%	35%	QMMGRAGR (SEC ID NO: 26)	QMMGRAGRP (SEC ID NO: 27)
GI:121723325	Hel308 Mba	Methanosarcina barkeri str. Fusaro	39%	60%	QMAGRAGR (SEC ID NO: 28)	QMAGRAGRP (SEC ID NO: 29)
GI:24418451	Hel308 Mac	Methanosarcina acetivorans	38%	60%	QMAGRAGR (SEC ID NO: 30)	QMAGRAGRP (SEC ID NO: 31)
GI:294495240	Hel308 Mmah	Methanohalophilus mahii DSM 5219	38%	60%	QMAGRAGR (SEC ID NO: 32)	QMAGRAGRP (SEC ID NO: 33)
GI:24418450	Hel308 Mmaz	Methanosarcina mazei	38%	60%	QMAGRAGR (SEC ID NO: 34)	QMAGRAGRP (SEC ID NO: 35)
GI:116665562	Hel308 Mth	Methanosaeta thermophila PT	39%	46%	QMAGRAGR (SEC ID NO: 36)	QMAGRAGRP (SEC ID NO: 37)
GI:336477283	Hel308 Mzh	Methanosalsum zhilinae DSM 4017	39%	57%	QMAGRAGR (SEC ID NO: 38)	QMAGRAGRP (SEC ID NO: 39)
GI:298674154	Hel308 Mev	Methanohalobium evestigatum Z-7303	38%	61%	QMAGRAGR (SEC ID NO: 40)	QMAGRAGRP (SEC ID NO: 41)
GI:500195255	Hel308	Methanococcus	36%	32%	QCIGRAGR (SEC	QCIGRAGRP (SEC

NO. GI	Helicasa	Nombres	% Identidad	% Identidad	Motivo Hel308	Motivo Hel308 ampliado
	Mma	maripaludis			ID NO: 42)	ID NO: 43)
GI:490388033	Hel308 Nma	Natrialba magadii	37%	43%	QMMGRAGR (SEC ID NO: 44)	QMMGRAGRP (SEC ID NO: 45)
GI:226740606	Hel308 Mbo	Methanoregula boonci 6A8	38%	45%	QMAGRAGR (SEC ID NO: 46)	QMAGRAGRP (SEC ID NO: 47)
GI:497573451	Hel308 Fac	Ferroplasma acidarmanus	34%	32%	QMIGRAGR (SEC ID NO: 48)	QMIGRAGRP (SEC ID NO: 49)
GI:256810263	Hel308 Mfe	Methanocaldococcus fervens AG86	40%	35%	QCIGRAGR (SEC ID NO: 50)	QCIGRAGRP (SEC ID NO: 51)
GI:18202572	Hel308 Mja	Methanocaldococcus jannaschii	24%	22%	QCIGRAGR (SEC ID NO: 52)	QCIGRAGRP (SEC ID NO: 53)
GI:502864579	Hel308 Min	Methanocaldococcus infernus	41%	33%	QCIGRAGR (SEC ID NO: 54)	QCIGRAGRP (SEC ID NO: 55)
GI:88603707	Hel308 Mhu	Methanospirillum hungatei JF-1	36%	40%	QMAGRAGR (SEC ID NO: 56)	QMAGRAGRP (SEC ID NO: 57)
GI:635552454	Hel308 Afu	Archaeoglobus fulgidus DSM 4304	40%	40%	QMAGRAGR (SEC ID NO: 58)	QMAGRAGRP (SEC ID NO: 59)
GI:502709689	Hel308 Htu	Haloterrigena turkmenica	35%	43%	QMAGRAGR (SEC ID NO: 60)	QMAGRAGRP (SEC ID NO: 61)
GI:495257384	Hel308 Hpa	Haladaptatus paucihalophilus DX253	38%	45%	QMFGRAGR (SEC ID NO: 62)	QMFGRAGRP (SEC ID NO: 63)
GI:15791161	Helicasa tipo Ski2	Halobacterium sp. NRC-1	36.8%	42.0%	QMFGRAGR (SEC ID NO: 64)	QMFGRAGRP (SEC ID NO: 65)
**: véase la Publicación Internacional No. WO 2013/057495 para más detalles de la secuencia de dicha helicasa Hel308.						

También se puede usar en las presentes realizaciones una variante o mutante de una helicasa Hel308 que retenga la unión de polinucleótidos y las actividades de la enzima helicasa. Dicha variante o mutante se puede obtener de acuerdo con métodos que son conocidos en la técnica, incluyendo mutagénesis específica del sitio del ácido nucleico que codifica la helicasa Hel308 nativa (Zoller, M. J., Curr. Opin. Biotechnol., 3: 348-354, (1992)).

Además, como se señaló anteriormente y como se conoce en la técnica, las helicasas Hel308 están en la familia SF2 y son helicasas de 3' a 5' (que también pueden denominarse helicasas de tipo A). Los dominios centrales de varias helicasas pueden incluir motivos comunes entre sí, como los pliegues de unión a RecA que contienen el motivo Walker A (que también se puede denominar motivo I) y el motivo Walker B (que también se puede denominar motivo II) involucrados en la unión de nucleótidos e hidrólisis, y el motivo VI. Para más detalles, véase Flechsig et al., "In Silico Investigation of Conformational Motions in Superfamily 2 Helicase Proteins", PLoS One: 6 (7): e 21809 (2011). Además, las helicasas de la familia SF2 pueden compartir nueve motivos conservados, que pueden denominarse Q, I, Ia, Ib, II, III, IV, V y VI. Debido a la secuencia del motivo II (DEAD (SEC ID NO: 2) o DEAH (SEC ID NO: 3) o DEXH), la familia de helicasas SF2 también puede denominarse helicasas de proteínas DEAD-box (SEC ID NO: 2) o box de DEAH (SEC ID NO: 3). Las helicasas incluidas en la familia SF2 incluyen la familia tipo RecQ y las enzimas tipo Snf2. Muchas helicasas SF2 son de tipo A, con algunas excepciones, como la familia XPD. Los estudios de cristalografía de rayos X de la familia SF2 sugieren que los motivos de helicasa conservados están estrechamente asociados con la estructura terciaria de la proteína, y que pueden formar un gran dominio funcional. Para más detalles, véase Tuteja et al., "Unraveling DNA Helicasas: Motif, structure, mechanism and function", European Journal of Biochemistry 271 (10): 1849-1863 (2004), y Hall et al., "Helicase motifs: the engine that powers DNA unwinding", Molecular Microbiology 34: 867-877 (1999). La Fig. 16, que está adaptada en Tuteja, ilustra esquemáticamente varios motivos que se han identificado en la familia SF2, por ejemplo, las helicasas DEAD-box (SEC ID NO: 2), de las cuales Hel308 es un miembro. Como se describe en Tuteja, las boxes abiertas representan motivos conservados. La secuencia de

consenso de cada motivo de helicasa está representada por códigos de una sola letra, por ejemplo, "C" en la FIG. 16 puede ser D, E, H, K o R; "O" en la FIG. 16 puede ser S o T; y "X" en la FIG. 16 puede ser cualquier aminoácido. Los nombres asignados a los motivos, por ejemplo, Q, I, Ia, Ib, II, III, IV, V y VI, también se muestran en la FIG. 16. Como se señaló más arriba, el motivo I se puede denominar motivo Walker A, y se menciona en Tuteja como ATPaseA Walker I, y el motivo II se puede denominar motivo Walker B y en Tuteja se denomina Walker ATPaseB II. Los números entre los motivos, a los que apuntan las flechas, son intervalos típicos de residuos de aminoácidos interpuestos entre los motivos.

Además, como se describe en el documento WO 2013/057495, una helicasa Hel308 puede incluir el(los) motivo(s) de aminoácidos Q-X1-X2-G-R-A-G-R (SEC ID NO: 66), en donde X1 puede ser C, M o L; X1 puede ser C; X2 puede ser cualquier residuo, incluido un residuo hidrófobo o neutro, como A, F, M, C, V, L, I, S, T, P o R. Opcionalmente, el terminal R en el(los) motivo(s) anterior(es) se puede acoplar a una P.

Dadas las enseñanzas y la guía proporcionadas en el presente documento, cualquier experto en la técnica podrá determinar si una helicasa de referencia es una helicasa Hel308 al determinar la identidad de secuencia o alineación con una o más de las helicasas Hel308 ejemplificadas anteriormente.

Además, dadas las enseñanzas y la guía proporcionadas en el presente documento, cualquier experto en la técnica podrá mutar adecuadamente la helicasa Hel308 para reducir la translocación fraccional de un polinucleótido a través de un poro, por ejemplo, al reducir la velocidad de hidrólisis que realiza la helicasa Hel308, mediante la mutación de un motivo del Hel308 análogamente como un motivo homólogo de otra proteína de una manera que pueda retardar la hidrólisis. Como ejemplo, Tanaka et al., "ATPase/helicase motif mutants of Escherichia coli PriA protein essential for recombination-dependent DNA replication", *Genes to Cells* 8: 251-261 (2003), describen mutantes de la proteína PriA (una helicasa tipo DEXH) que llevan sustituciones de aminoácidos en sus motivos de helicasa ATPasa/ADN conservados, a saber, los motivos Walker A, B y QXXGRXGR. Según Tanaka, ciertos mutantes estaban altamente comprometidos en la hidrólisis de ATP en ciertas condiciones, y todas las proteínas mutantes Walker A y Walker B mostraron actividad de helicasa de ADN altamente atenuada en ciertas condiciones. Por consiguiente, se puede esperar que las mutaciones en los motivos de Walker A y Walker B de una helicasa Hel308 que son análogas a las descritas en Tanaka pueden atenuar la actividad de la helicasa del ADN o retardar la hidrólisis de ATP, y se puede esperar que frenen la translocación fraccional del polinucleótido a través del poro y, por lo tanto, que mejoren la caracterización de ese polinucleótido. Como otro ejemplo, Hishida et al., "Role of Walker Motif A of RuvB Protein in Promoting Branch Migration of Holliday Junctions: Walker motif A mutations affect ATP binding, ATP hydrolyzing, and DNA binding activities of RuvB", *Journal of Biological Chemistry*, 274 (36): 25335-25342 (1999) describe mutantes de la proteína RuvB de Escherichia coli, una helicasa de ADN hexamérica dependiente de ATP. Según Hishida, ciertas mutaciones puntuales en el motivo A de Walker afectaron las actividades de RuvB de la hidrólisis del ATP y la unión del ATP, así como las de la unión al ADN, la formación de hexámeros y la promoción de la migración de la rama. Por consiguiente, puede esperarse que las mutaciones en los motivos de Walker A de una helicasa Hel308 que son análogas a las descritas en Hishida puedan afectar a la hidrólisis de ATP y la unión de ATP, que puede esperarse que reduzcan la translocación fraccional de un polinucleótido a través de un poro y, por lo tanto, pueden mejorar la caracterización de ese polinucleótido en ciertas realizaciones.

Por consiguiente, la descripción proporciona un método para caracterizar un polinucleótido diana. El método puede incluir (a) aplicar una diferencia de potencial a través de un poro en contacto con una helicasa Hel308 y un polinucleótido diana; (b) medir una o más señales producidas por una o más etapas de translocación fraccional por dicha helicasa Hel308 de dicho polinucleótido diana a través de dicho poro; y (c) caracterizar dicho polinucleótido diana a partir de dicha una o más señales producidas por dichas etapas de translocación fraccional.

La descripción proporciona además un método para caracterizar un polinucleótido diana en el que la diferencia de potencial comprende una diferencia de potencial eléctrico. También se proporciona un método para caracterizar un polinucleótido diana en el que la señal incluye una señal eléctrica o una señal óptica. La señal eléctrica puede ser una medida seleccionada entre la corriente, el voltaje, el efecto túnel, la resistencia, el potencial, el voltaje, la conductancia; la medición eléctrica y transversal. La señal eléctrica incluye una corriente eléctrica que pasa a través del poro.

En otros aspectos, la descripción proporciona un método para caracterizar un polinucleótido diana en el que la etapa de translocación fraccional incluye una primera etapa de translocación fraccional de un ciclo de translocación completo de la helicasa Hel308. La etapa de translocación fraccional también puede incluir una segunda etapa de translocación fraccional de un ciclo de translocación completo de la helicasa Hel308. La translocación del polinucleótido diana puede ser en una dirección opuesta a la fuerza aplicada en el polinucleótido que se transloca a través del poro o en una dirección con la fuerza aplicada en el polinucleótido que se transloca a través del poro.

Además, se proporciona un método para caracterizar un polinucleótido diana en el que uno o más residuos de nucleótidos en el polinucleótido diana se caracterizan utilizando señales eléctricas obtenidas de dos etapas fraccionarias de un ciclo de translocación completo con una precisión superior al 50% en comparación con la caracterización de uno o más nucleótidos utilizando una única señal eléctrica obtenida de un ciclo completo de translocación.

Además, se proporciona un método para caracterizar un polinucleótido diana en el que el poro es un poro biológico.

El poro biológico puede ser un poro polipeptídico o un poro polinucleotídico. En algunos aspectos, el poro del polipéptido tiene una zona de constricción de cinco nucleótidos o menos. En otros aspectos, el poro del polipéptido incluye una porina A de *Mycobacterium smegmatis* (MspA). La MspA puede tener una secuencia de aminoácidos de SEC ID NO: 1 o tener al menos 15%, al menos 20%, al menos 25%, al menos 30%, al menos 35%, al menos 40%, al menos 45%, al menos 50%, al menos 55%, al menos 60%, al menos 65% o al menos 70% de homología con la SEC ID NO: 1.

También se proporciona un método para caracterizar un polinucleótido diana en el que el poro es un poro en estado sólido o un poro híbrido biológico y en estado sólido. El poro híbrido en estado sólido y biológico incluye un poro híbrido en estado sólido polipéptido o un poro híbrido en estado sólido polinucleótido.

La descripción proporciona además un método para caracterizar un polinucleótido diana en el que la helicasa Hel308 es una de las helicasas mostradas en las Tablas 1 y 2 o una variante de las mismas. Además, se proporciona un método para caracterizar un polinucleótido diana en el que el polinucleótido diana se selecciona del grupo que consiste en un polinucleótido monocatenario, bicatenario y parcialmente bicatenario.

En algunas realizaciones, la caracterización del polinucleótido a partir de dichas, al menos dos, señales producidas por dichas etapas de translocaciones fraccionales comprende la aplicación de un algoritmo de Viterbi modificado.

En algunas realizaciones, el método incluye además (d) después de la etapa (c), variar al menos un parámetro para variar la sincronización de una o más etapas de translocación fraccional por dicha helicasa Hel308 de dicho polinucleótido diana a través de dicho poro; y (e) repetir las etapas (a) - (c) utilizando el, al menos un, parámetro variado. El método puede incluir además combinar las señales producidas durante las etapas (c) y (e) y caracterizar dicho polinucleótido diana en base a las señales combinadas. En algunas realizaciones, el, al menos un, parámetro variado se selecciona del grupo que consiste en temperatura, concentración de sal, concentración de cofactor, concentración de producto de ATP (como pirofosfato inorgánico), concentración de ADP, pH y la helicasa Hel308 particular utilizada.

En algunas realizaciones, dicha caracterización comprende detectar e identificar niveles en una o más señales y determinar y emitir una secuencia del polinucleótido diana en función de los niveles detectados e identificados.

De manera ilustrativa, dicha detección e identificación de los niveles en una o más señales incluye la salida de uno o más de los niveles completos, los niveles fraccionales, todos los niveles y los identificadores de nivel.

Dicha determinación y salida de la secuencia del polinucleótido diana en función de los niveles detectados e identificados puede incluir tomar como entrada uno o más de los niveles completos, niveles fraccionales, todos los niveles e identificadores de nivel, resolviendo una pluralidad de secuencias basadas en dicha entrada, y seleccionando y resolviendo al menos una de las secuencias resueltas en función de la información de confianza respecto a la pluralidad de secuencias resueltas.

Dicha determinación y resolución de la secuencia del polinucleótido diana en función de los niveles detectados e identificados puede incluir tomar como entrada uno o más de los niveles completos, niveles fraccionales, todos los niveles e identificadores de nivel, resolviendo una pluralidad de secuencias basadas en dicha entrada, y seleccionando y concatenando entre sí porciones de una pluralidad de las secuencias resueltas basadas en información de confianza respecto a las porciones de la pluralidad de secuencias resueltas.

Dicha determinación y resolución de la secuencia del polinucleótido diana en función de los niveles detectados e identificados puede incluir tomar como entrada uno o más de los niveles completos, niveles fraccionales, todos los niveles e identificadores de nivel, resolviendo una pluralidad de secuencias basadas en dicha entrada, comparando las secuencias resueltas para modelar las secuencias, y seleccionando y resolviendo al menos una de las secuencias resueltas en función de la información de confianza respecto a la comparación de la secuencia resuelta con la secuencia modelo.

Dicha determinación y resolución de la secuencia del polinucleótido diana en función de los niveles detectados e identificados puede incluir tomar como entrada uno o más de los niveles completos, niveles fraccionales, todos los niveles e identificadores de nivel, resolviendo una pluralidad de secuencias basadas en dicha entrada, comparando las secuencias resueltas para modelar secuencias, y seleccionando y concatenando entre sí porciones de una pluralidad de las secuencias resueltas basándose en la información de confianza respecto a la comparación de porciones de la pluralidad de secuencias resueltas con la secuencia modelo.

La presente descripción también proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana a través de un poro. El método puede incluir: (a) aplicar una diferencia de potencial a través de un poro en contacto con una helicasa Hel308 y un polinucleótido diana; (b) poner en contacto la helicasa Hel308 con una concentración de un sustrato de helicasa Hel308 que es diferente de una concentración de referencia del sustrato, la concentración del sustrato produce un cambio en la duración de una etapa de translocación fraccional proporcional a una diferencia en la concentración del sustrato en comparación con la concentración de referencia, y (c) medir una señal producida por una etapa de translocación fraccional del polinucleótido diana a través del poro. La etapa (b) también puede incluir el uso de un análogo o inhibidor de sustrato para lograr un cambio en la duración de una etapa

de translocación fraccional. Por consiguiente, cualquiera de los análogos de sustrato o inhibidores de nucleótidos descritos en el presente documento o conocidos en la técnica se pueden usar en el método de la descripción para modular una etapa de translocación fraccional como un sustrato de helicasa Hel308, un sustrato Hel308 que se está utilizando como una concentración de referencia o tanto un sustrato de helicasa Hel308 como un sustrato de Hel308 que se está utilizando como concentración de referencia.

Un sustrato de helicasa Hel308 que es capaz de modular la etapa de translocación fraccional de un polinucleótido puede ser un nucleótido o un análogo de nucleótido que puede ser hidrolizado por la helicasa. El sustrato de nucleótido proporciona energía para desenrollar un polinucleótido de cadena doble o parcialmente de cadena doble o para desplazar un polinucleótido de cadena sencilla a través de un poro. Un sustrato común para una helicasa Hel308 incluye, por ejemplo, ATP. Los sustratos de helicasa Hel308 también incluyen nucleótidos y análogos de nucleótidos que pueden hidrolizarse mediante helicasa.

Como se describe en este documento, el tiempo de permanencia para una o más etapas de translocación fraccional que están relacionadas con la unión al sustrato de nucleótido puede ser inversamente proporcional a la concentración del sustrato de helicasa Hel308. Por ejemplo, bajo ciertas condiciones ensayadas, el tiempo de permanencia para solo una de las dos etapas de translocación fraccional observadas por translocación de nucleótidos es inversamente proporcional a la concentración del sustrato de helicasa Hel308. Por lo tanto, una etapa de translocación puede ser sensible a la concentración de sustrato mientras que la otra etapa de translocación no lo es.

La modulación de una etapa de translocación fraccional para obtener una longitud diferente de las etapas de translocación fraccional se puede lograr cambiando la concentración del sustrato de helicasa Hel308. El grado o la magnitud de la modulación se puede determinar de modo que cualquier experto en la técnica pueda seleccionar una longitud particular de etapas de translocación fraccional adecuada para alguna caracterización de polinucleótidos diana deseados. El grado de modulación puede determinarse colocando una helicasa Hel308 en una concentración de un sustrato de Helicasa Hel308 que sea diferente de una concentración de referencia del sustrato. El cambio en la concentración del sustrato en comparación con la concentración de referencia da como resultado un tiempo de permanencia diferente de una etapa de translocación fraccional que es proporcional a la diferencia en la concentración del sustrato en comparación con la concentración de referencia.

Por consiguiente, la etapa de translocación fraccional de un polinucleótido diana a través de un poro puede modularse utilizando una concentración de un sustrato de helicasa Hel308 que sea diferente de una concentración de referencia del sustrato. Otros componentes dentro de la solución de helicasa o las condiciones de reacción también se pueden usar para alterar el tiempo de permanencia de una etapa de translocación fraccional y, por lo tanto, la longitud de las etapas de translocación fraccional para un solo ciclo de translocación. Los diferentes etapas de translocación fraccional pueden usarse de manera similar para adquirir más información de la señal para aumentar la precisión de la caracterización del polinucleótido diana.

Los componentes de la reacción y las condiciones de reacción que influyen, por ejemplo, en la cinética de la unión del sustrato a la helicasa Hel308 y la hidrólisis del sustrato por la helicasa se pueden usar para alterar el tiempo de permanencia de una etapa de translocación fraccional. Tales otros factores incluyen, por ejemplo, temperatura, concentración de metal, incluyendo la concentración del metal divalente, concentración del ión, viscosidad del disolvente de las condiciones de reacción. La etapa de hidrólisis puede verse influida, por ejemplo, por los factores y condiciones anteriores, así como por la concentración de fosfato y/o pirofosfato. Además, el voltaje a través del poro puede, por ejemplo, afectar la unión del sustrato y/o la pausa de helicasa que constituye el tiempo de permanencia de una helicasa Hel308. Otros factores incluyen, por ejemplo, el pH, el tipo de catión o la concentración y el tipo de catión divalente, las mutaciones de helicasa, etc., y todos pueden afectar a los tiempos de permanencia. En este sentido, por ejemplo, se puede utilizar una concentración creciente de pirofosfato para disminuir la velocidad catalítica de la helicasa Hel308 y, por lo tanto, aumentar el tiempo de permanencia. Además, por ejemplo, el ortovanadato de sodio y el hidrato de la sal de adenosina 5'-(β,γ -imido)trifosfato de litio también se pueden usar para disminuir la actividad de la helicasa. El uso de pirofosfato y análogos de nucleótidos para modular la actividad helicasa se ilustra a continuación en el Ejemplo V.

A medida que aumentan las diferencias de corriente entre las etapas secuenciales, también aumentan los beneficios de usar estados fraccionales para el análisis de datos. En la primera aproximación, las etapas de translocación fraccional tomarán los valores que se encuentren entre las etapas de translocación completas adyacentes. Si una etapa de translocación fraccional es mucho menor que 1/2 nucleótidos (0,3 Angstroms), el valor fraccional puede ser, en algunos casos, o incluso en muchos casos, difícil o incluso imposible de observar. Si la etapa de translocación fraccional tiene exactamente la longitud de 1/2 nucleótido, entonces la corriente resultante puede ser, en promedio, lo más distinta posible de los valores de corriente anteriores y posteriores que corresponden a pasos de nucleótidos completos. La modificación de la enzima puede permitir el reposicionamiento de subunidades de polímeros mediante fracciones de un nanómetro. Esto puede suceder a través de modificaciones enzimáticas que aumenten o disminuyan la altura relativa del sitio de hidrólisis activa de la enzima, hasta la constricción limitante del nanoporo. En algunas realizaciones, esto se puede lograr mediante la adición o eliminación de aminoácidos de helicasa, o la sustitución de aminoácidos con un radio hidrodinámico más grande. En otras realizaciones, esto puede lograrse mediante la alteración de la carga de aminoácidos que puede alterar la repulsión o atracción electrostática al borde del nanoporo. Sin desear estar limitado por ninguna teoría, si la hipótesis "basada en la sujeción" es correcta (como se describe con

mayor detalle con referencia a la Figura 3), puede ser posible que una mutación particular afecte al grado en que la helicasa empuja hacia arriba el complejo helicasa-polinucleótido, que puede traducirse en un cambio en el porcentaje de translocación del eje z del nucleótido.

Se contempla ajustar la duración de las etapas de translocación fraccional: es razonable esperar que ciertas mutaciones en el dominio ATPasa de la helicasa afecten a la velocidad en la que se hidroliza el ATP. Esto, a su vez, se espera que afecte el tiempo de permanencia para cualquiera de las etapas de translocación fraccional. Por ejemplo, si la velocidad de hidrólisis se redujo, se espera que aumente el tiempo de permanencia para cualquiera de las etapas de translocación fraccional. Otras mutaciones podrían afectar a la velocidad a la que ATP se une a la helicasa (k_{on}). En este caso, a medida que aumenta el tiempo que tarda un ATP en unirse, aumentará el tiempo de permanencia para la etapa de translocación fraccional correspondiente.

Una concentración de referencia de una helicasa Hel308 puede ser, por ejemplo, la cantidad de sustrato generalmente utilizado en una caracterización de polinucleótidos diana o puede ser diferente. Por ejemplo, si la concentración de un sustrato de helicasa Hel308 que se usa generalmente es de 1,0 mM, entonces 1 mM correspondería a la concentración de referencia. La concentración de referencia puede derivarse empíricamente u obtenerse de informes conocidos en la técnica. En este ejemplo específico, la concentración de sustrato diferente a 1 mM sería el sustrato de helicasa Hel308 que es diferente de la concentración de referencia. Como se describe más adelante, se pueden emplear diversas concentraciones de un sustrato de helicasa Hel308 y un sustrato de referencia para modular o determinar la cantidad de alteración de una etapa de translocación fraccional.

La concentración de sustrato de helicasa Hel308 y la concentración de sustrato de referencia pueden variar siempre que ambas concentraciones no sean concentraciones de saturación. Ilustrativamente, una concentración de saturación de un sustrato de helicasa Hel308 es aproximadamente 1 mM de sustrato de nucleótido. Por consiguiente, si una concentración de referencia es 1 mM, entonces la concentración del sustrato de helicasa Hel308 que se va a variar puede ser cualquier concentración inferior a 1 mM, incluyendo, por ejemplo, 0,1 μ M, 1,0 μ M, 10 μ M, 100 μ M, 200 μ M, 300 μ M, 400 μ M, 500 μ M, 600 μ M, 700 μ M, 800 μ M, 900 μ M. Dependiendo de la concentración de sustrato de helicasa Hel308 y/o la referencia, otras concentraciones ejemplares pueden ser, por ejemplo, 1,0 mM, 2,0 mM, 3,0 mM, 4,0 mM y 4,9 mM o menos. De manera similar, la concentración tanto del sustrato de helicasa Hel308 como la concentración del sustrato de referencia pueden ser concentraciones que no se sean saturantes, siempre que difieran. Por consiguiente, el sustrato de helicasa Hel308 y las concentraciones de referencia pueden ser cualquiera de las concentraciones ejemplares enumeradas anteriormente, así como cualquier concentración entre, por ejemplo, cualquier concentración que varíe entre 0,01 μ M y 5 mM y todas las concentraciones dentro de este intervalo.

Los métodos de la presente descripción para modular una etapa de translocación fraccional pueden realizarse como se describió previamente con respecto a los métodos para caracterizar un polinucleótido diana. Una vez que se determina una concentración de sustrato de helicasa Hel308 que sea adecuada para una necesidad particular, se puede emplear la concentración de sustrato en los métodos descritos en el presente documento para caracterizar el polinucleótido diana. De manera similar, se pueden realizar determinaciones similares con componentes y condiciones de la reacción que influyen, por ejemplo, en la cinética de la unión del sustrato y la hidrólisis para determinar la concentración del componente o las condiciones de reacción adecuadas para cualquier necesidad particular. Esa concentración o condición adecuada puede emplearse luego en los métodos de la presente descripción para caracterizar el polinucleótido diana. La nueva concentración de sustrato, la concentración del componente de reacción y/o la condición de reacción darán como resultado un tiempo de permanencia diferente que puede proporcionar información de señal adicional para mejorar la precisión de la determinación, de una manera tal como se describe a continuación con referencia al Ejemplo IX.

Por consiguiente, la descripción proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana que incluye además caracterizar el polinucleótido diana a partir de una o más señales de la una o más etapas de translocación fraccional. La caracterización puede incluir la identificación de uno o más de: (1) la secuencia del polinucleótido diana; (2) la modificación del polinucleótido diana; (3) la longitud del polinucleótido diana; (4) la identidad del polinucleótido diana; (5) la fuente del polinucleótido diana, o (6) la estructura secundaria del polinucleótido diana.

La descripción también proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana en el que el método emplea una diferencia de potencial que incluye una diferencia de potencial eléctrico. Además, se proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana en el que la señal producida por la etapa de translocación fraccional incluye una señal eléctrica o una señal óptica. Adicionalmente, se proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana. La señal eléctrica es una medida seleccionada entre corriente, voltaje, efecto túnel, resistencia, potencial, voltaje, conductancia; medición eléctrica y transversal. La señal eléctrica también puede ser una corriente eléctrica que pase a través del poro.

También se proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana en el que la concentración de sustrato sea una concentración subsaturante del sustrato de helicasa Hel308. En algunas realizaciones, la concentración de referencia es una concentración de saturación del sustrato de helicasa Hel308. En otros aspectos, tanto la concentración de sustrato como la concentración de referencia son concentraciones

subsaturantes del sustrato de helicasa Hel308. Además, se proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana en el que el sustrato de helicasa Hel308 es el trifosfato de adenosina (ATP).

5 Aún más, se proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana en el que la etapa de translocación fraccional incluye una primera etapa de translocación fraccional de un ciclo de translocación completo de la helicasa Hel308 o una segunda etapa de translocación fraccional de un ciclo de translocación completo de la helicasa Hel308. Una translocación del polinucleótido diana puede ser en una dirección opuesta a la fuerza aplicada en el polinucleótido que se transloca a través del poro o en una dirección con la fuerza aplicada en el polinucleótido que se transloca a través del poro.

10 La descripción también proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana en el que uno o más residuos de nucleótidos en el polinucleótido diana se caracterizan utilizando señales eléctricas obtenidas de dos etapas fraccionarias de un ciclo completo de translocación con una precisión mayor que el 50% en comparación con la caracterización de uno o más nucleótidos utilizando una única señal eléctrica obtenida de un ciclo de translocación completo. En algunos aspectos del método de la descripción, uno o más residuos de nucleótidos en el polinucleótido diana se caracterizan con una mayor precisión a una concentración de sustrato más baja en comparación con la concentración de referencia.

15 Adicionalmente, se proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana en el que el poro es un poro biológico. El poro biológico puede ser un poro polipeptídico o un poro polinucleotídico. En algunos aspectos, el poro del polipéptido tiene una zona de constricción de cinco nucleótidos o menos. En otros aspectos, el poro del polipéptido incluye una porina A de *Mycobacterium smegmatis* (MspA). La MspA puede tener una secuencia de aminoácidos de SEC ID NO: 1 o tener al menos 15%, al menos 20%, al menos 25%, al menos 30%, al menos 35%, al menos 40%, al menos 45%, al menos el 50%, al menos el 55%, al menos el 60%, al menos el 65%, al menos el 70%, al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, o al menos un 99% de homología con la SEC ID NO: 1.

25 Se proporciona además un método para modular una etapa de translocación fraccional de un polipéptido diana en el que el poro es un poro en estado sólido o un poro híbrido en estado sólido y biológico. El poro híbrido en estado sólido y biológico puede ser un poro híbrido en estado sólido polipéptido o un poro híbrido en estado sólido polinucleotídico.

También se proporciona un método para modular una etapa de translocación fraccional de un polinucleótido diana en el que la helicasa Hel308 en el método incluye la helicasa que se muestra en las Tablas 1 y 2 o una variante de la misma. El polinucleótido diana se selecciona del grupo que consiste en un polinucleótido monocatenario, bicatenario y parcialmente bicatenario.

30 La presente descripción proporciona además una composición para caracterizar un polinucleótido diana. La composición incluye un poro, una helicasa Hel308 y un polinucleótido diana contenido en una solución de ATP menos de 1 mM o una solución de un análogo de nucleótido. En algunos aspectos de la composición, la solución de menos de 1 mM de ATP tiene 0,1 μ M, 1,0 μ M, 10 μ M, 100 μ M, 0,5 mM o 0,9 mM de ATP.

35 Una composición de la presente descripción puede incluir cualquiera de los componentes descritos anteriormente o a continuación que se usan en los métodos de la presente descripción para caracterizar cualquier polinucleótido o para modular una etapa de translocación fraccional de una translocación de polinucleótido diana. Por ejemplo, la composición puede incluir un poro como se describió anteriormente. De acuerdo con las enseñanzas y las orientaciones proporcionadas en el presente documento, el poro puede ser, por ejemplo, un poro biológico tal como un poro polipeptídico o poro de polinucleótido. Alternativamente, el poro puede ser un poro en estado sólido o un poro híbrido como se describió anteriormente.

40 Además, la composición incluirá un polinucleótido diana para la caracterización, una helicasa Hel308 y un sustrato de helicasa Hel308. Al igual que con el poro, el polinucleótido diana, la helicasa Hel308 y el sustrato de helicasa Hel308 pueden ser cualquiera de los polinucleótidos, helicasas Hel308, sustratos y variantes y análogos ejemplares descritos en este documento, así como los conocidos en la técnica.

45 Por consiguiente, la descripción proporciona una composición para caracterizar un polinucleótido diana en el que el poro es un poro biológico. El poro biológico puede ser un poro polipeptídico o un poro polinucleotídico. El poro del polipéptido puede tener una zona de constricción de cinco nucleótidos o menos y puede ser una porina A (MspA) de *Mycobacterium smegmatis*. La MspA puede tener una secuencia de aminoácidos de SEC ID NO: 1 o tener al menos 15%, al menos 20%, al menos 25%, al menos 30%, al menos 35%, al menos 40%, al menos 45%, al menos el 50%, al menos el 55%, al menos el 60%, al menos el 65%, al menos el 70%, al menos el 75%, al menos el 80%, al menos el 85%, al menos el 90%, al menos el 95%, o al menos un 99% de homología con la SEC ID NO: 1.

50 También se proporciona una composición para caracterizar un polinucleótido diana en el que el poro es un poro en estado sólido. Adicionalmente, se proporciona una composición para caracterizar un polinucleótido diana en el que el poro es un poro híbrido biológico y en estado sólido. El poro híbrido en estado sólido y biológico puede ser un poro híbrido en estado sólido polipéptido o un poro híbrido en estado sólido polinucleotídico.

Además, se proporciona una composición para caracterizar un polinucleótido diana en el que la helicasa Hel308 es una helicasa mostrada en las Tablas 1 y 2 o una variante de las mismas. Adicionalmente, se proporciona una composición para caracterizar un polinucleótido diana en el que el polinucleótido diana se selecciona del grupo que consiste en un polinucleótido monocatenario, bicatenario y parcialmente bicatenario.

- 5 Por consiguiente, los siguientes ejemplos tienen la intención de ilustrar, pero no limitar, la presente descripción.

Ejemplo I

Etapas de translocación fraccional con una Helicasa Hel308

El Ejemplo I describe las etapas de translocación fraccional observadas con una helicasa Hel308 ejemplar.

10 Se formaron bicapas lipídicas a partir de 1,2-difitanoil-sn-glicero-3-fosfolina (Avanti Polar Lipids). La bicapa abarcaba una abertura horizontal de aprox. 20 micras de diámetro en teflón. Se agregó M2-NNN-MspA al lado conectado a tierra de la bicapa a una concentración de ~2,5 ng/ml. Una vez que se insertó un solo poro, el compartimiento se lavó con un tampón experimental para evitar nuevas inserciones. Un amplificador de pinza de parche Axopatch-200B (Axon Instruments) aplicó un voltaje a través de la bicapa de 180 mV y midió las corrientes iónicas. La señal analógica se filtró a paso bajo a 50 kHz con un filtro Bessel de 4 polos y luego se digitalizó a cinco veces la frecuencia del filtro a 15 paso bajo. La adquisición de datos se controló con un software personalizado escrito en LabWindows/CVI (National Instruments). Los compartimentos de ~60 µl en ambos lados de la bicapa contenían un tampón experimental de KCl 0,3 M, EDTA 1 mM, DTT 1 mM, MgCl₂ 10 mM y HEPES/KOH 10 mM tamponados a pH 8,0. Como motor se usó Hel 308 Tga natural o Phi29 natural. En presencia de Hel308 Tga, el tampón se complementó con ATP 1 mM. En presencia de Phi29, el tampón se complementó con 100 µM de dCTP, dATP, dTTP y dGTP.

20 Las figuras 2A-2C muestran la comparación de eventos de translocación de la polimerasa Phi29 y helicasa Hel308 Tga, de acuerdo con algunas realizaciones. La Fig. 2A muestra las etapas de translocación fraccional observadas con una helicasa Hel308 Tga en comparación con las etapas de translocación observadas con una ADN polimerasa phi29 (DNAP). El polinucleótido de translocación (SEC ID NO: 67:5Phos/AAACCTTCCXCCCGTACCGTGCCGTACCGTTCCGTTCCGTACCGTATTTTTTTT
25 TCTCACTATCGCATTCTCATGCAGGTCGTAGCC donde X = abásico) se hibridó a un polinucleótido que contenía colesterol (SEC ID NO: 68: CCTGCATGAGAATGCGATAGTGAGA TTTTTTTTTTTTTTTTTTTT-CholTEG). Se utilizó el nanoporo MspA-M2. La cantidad de niveles observados para la translocación del polinucleótido helicasa Hel308 Tga fue casi el doble de la cantidad de niveles observados para la DNAP de phi29. Las líneas dibujadas entre trazos indican niveles correspondientes. La traza phi29 (arriba) es un consenso, mientras que la traza helicasa Hel308 (abajo) es un evento de translocación individual medido. Un consenso puede referirse a la combinación de niveles detectados de manera confiable a partir de múltiples lecturas de la misma secuencia. Una combinación de este tipo puede ser potencialmente más confiable que una sola lectura, ya que puede no incluir necesariamente errores como los que pueden ocurrir con la translocación de una sola molécula, por ejemplo, la "omisión" de nucleótidos o la "alternancia" de nucleótidos, como se conoce en la técnica.

35 La Fig. 2B muestra las etapas de translocación fraccional observadas con una helicasa Hel308 Tga en comparación con los niveles de corriente pronosticados generados por una plantilla de polinucleótidos de hebra única que se translocan a través de un nanoporo MspA-M2 utilizando una polimerasa Phi29 como el motor molecular con los observados utilizando una helicasa Tga Hel308 como motor molecular. El polinucleótido de translocación (SEC ID NO: 69:5Phos/CTCACCTATCCTTCCACTXXCCCCCTTTGGGTTTAAATTTTTTCAGATCTCACTATC
40 TTTTAAAGTTTTCTCACTATCGCATTCTCATGCAGGTCGTAGCC donde X = abásico) se hibridó a un polinucleótido que contiene colesterol (SEC ID NO: 68: CCTGCATGAGAATGCGATAGTGAGATTTTTTTTTTTTTTTTTTTT-CholTEG). Se predice la traza phi29 (arriba), mientras que la traza helicasa Hel308 (abajo) es un evento medido de translocación. Una predicción, como un consenso, puede no incluir necesariamente errores como los que pueden ocurrir con la translocación de una sola molécula, por ejemplo, la "omisión" de nucleótidos o la "alternancia" de nucleótidos, como se conoce en la técnica. Una predicción puede referirse a datos simulados basados en una tabla k-mer, que se había recopilado previamente. El patrón predicho de Phi29 se basa en una etapa completa por base, e ilustra un tipo de patrón que puede esperarse con un motor molecular de etapa completo. En comparación, se puede ver claramente que Hel308 Tga helicasa tiene etapas fraccionarias.

50 La Fig. 2C muestra las etapas de translocación fraccional observadas con una helicasa Hel308 Tga, en comparación con las etapas de translocación observadas con un DNAP phi29. La secuencia de polinucleótidos de translocación es: SEC ID NO: 70:5Phos/CATCATCATCATCATXXCCCCCTAAACAAGAATAACCACGACTAGCATTTT TCAGATCTCACTATCGCATTCTCATATCCAGGTCGTAGCC. El polinucleótido de translocación se hibridó a un polinucleótido que contenía colesterol (SEC ID N°: 68: CCTGCATGAGAATGCGATAGTGAGA TTTTTTTTTTTTTTTTTTTT-CholTEG) y se pasó a través del nanoporo MspA-M2. La secuencia repetida simple 5'-
55 CAT-3' se usó para mostrar un patrón repetido. El número de niveles observados para la translocación del polinucleótido helicasa Hel308 Tga fue el doble del número de niveles observados para DNAP de phi29. Tanto los rastros de helicasa phi29 como Hel308 son trazas de consenso. El uso de trazas de consenso puede facilitar la comparación de los tamaños de las etapas de translocación entre diferentes motores moleculares, y puede reducir o eliminar artefactos como omisiones y alternancias que, de otro modo, pueden complicar la interpretación.

Sin desear estar limitado por ninguna teoría, al explicar con más detalle las etapas de translocación fraccional, se propone un mecanismo "basado en la sujeción". La Fig. 3 muestra un mecanismo propuesto "basado en la sujeción" para una etapa de translocación fraccional, de acuerdo con algunas realizaciones. El polinucleótido (línea negra rellena) está unido por la helicasa (forma con relleno de línea horizontal). Tras la unión de ATP (Etapa 1), la helicasa sufre un cambio conformacional (Etapa 2). Dado que el polinucleótido es sujetado por la helicasa, la ubicación del polinucleótido con respecto a la helicasa no cambia necesariamente. Un punto de referencia en la helicasa (triángulo gris) no se mueve con respecto al polinucleótido sujetado por la helicasa (ver punto de referencia en el polinucleótido sujetado, cuadrado gris). El cambio de conformación de la helicasa empuja el complejo de helicasa-polinucleótido fuera de la parte superior del nanoporo, tirando del polinucleótido en la constricción del poro (línea negra apuntada por una línea negra con punta de flecha) junto con ella. Un segundo punto de referencia de polinucleótido (círculo blanco) muestra el polinucleótido moviéndose con respecto a la constricción del poro durante el cambio conformacional (Etapa 2), lo que da como resultado un cambio de la corriente medida para la etapa fraccional. Finalmente, el ATP se hidroliza y la helicasa se transloca a lo largo del polinucleótido (Etapa 3). Esto hace que el polinucleótido mueva un nucleótido completo con respecto a la helicasa y el poro. En resumen, en la primera etapa de translocación fraccional, la helicasa Hel308 se une al ATP y experimenta un cambio conformacional que tira del polinucleótido sujetado por la helicasa y desplaza el polinucleótido por una fracción de un nucleótido, que a su vez produce un cambio de corriente mensurable. En la segunda etapa de translocación fraccional, el ATP se hidroliza y la helicasa Hel308 completa la translocación del nucleótido a través del nanoporo. Otros mecanismos pueden usarse adecuadamente para explicar las observaciones de corriente de las etapas de translocación fraccional.

Ejemplo II

La relación entre las concentraciones de ATP y las etapas de translocación fraccional

El Ejemplo II describe el efecto de la concentración de ATP en los tiempos de permanencia de las etapas de translocación fraccional.

Para aclarar más el mecanismo bioquímico de la etapa de translocación fraccional, se examinaron los tiempos de permanencia de las etapas de translocación fraccional con concentraciones variables de ATP. Los pocillos cis y trans se llenaron primero con una solución tampón que consistía en KCl 400 mM, HEPES 10 mM, pH 8. Se formó una bicapa lipídica que consistía en DPhPC pintando una mezcla de hexadecano y lípido sobre un poro de teflón de aprox. 25 μm de diámetro y se realizaron mediciones de conductancia para asegurar un sello Gigaohm entre la bicapa lipídica y el poro de teflón. Todas las mediciones eléctricas se realizaron utilizando un amplificador de pinza de parche Axopatch 200B conectado a un par de electrodos de Ag/AgCl conectados a los pocillos cis y trans. Después de la formación de la membrana, se inyectaron nanoporos MspA en el pocillo cis donde se monitorizó la incorporación de nanoporos en la bicapa lipídica a través de mediciones de conductancia. Tras la incorporación de un único nanoporo en la bicapa, la cámara cis se perfundió para evitar la inserción de múltiples poros. Luego se inyectó polinucleótido monocatenario en la cámara cis a una concentración final de 10 nM, se aplicó voltaje a través de la membrana y se detectó la translocación de polinucleótidos a través de poros por medio de respuestas de corriente transitoria. Tras la detección de la translocación de polinucleótidos, el voltaje se ajustó a 0 V y se inyectaron MgCl_2 1 mM, helicasa Hel308 115 nM y varias concentraciones de ATP (10 μM , 30 μM , 100 μM y 1 mM) en el pocillo cis. Luego se ajustó el voltaje a un potencial de mantenimiento (140 mV para ATP 0,01, 0,1 y 1 mM; 180 mV para ATP 0,03 mM) y se registraron las corrientes. El polinucleótido de translocación (SEC ID NO: 71:5Phos/CATCATCATCATCATCATXXCCCCCTAAACAAGAATACCGACTAGCATTTTTCAGATCTCACTATCGCATTCATGCAGGTCGTAGCC donde X = abásico) se hibridó a un polinucleótido que contenía colesterol (SEC ID NO: 68: CCTGCATGAGATGCGATAGTGAGA TTTTTTTTTTTTTTTTTTTT-CholTEG) antes de la inyección en el pocillo cis así como se muestra en las figs. 14A-14D (descrito con mayor detalle en otra parte del presente documento). De esta manera, el extremo 5' del polinucleótido se translocó primero a través del nanoporo y luego se retiró a través del nanoporo mediante el procesamiento de la helicasa Hel308. El amplificador Axopatch registró la respuesta de corriente del sistema a una frecuencia de muestreo de 50 kHz y con un filtro a paso bajo de 10 kHz. Las transiciones de etapa debidas al procesamiento de helicasa Hel308 del polinucleótido a través del nanoporo, incluidas las etapas de translocación fraccional, fueron claramente identificables dentro de este intervalo de frecuencia. Tras la experimentación, se utilizaron algoritmos informáticos para identificar eventos de translocación de polinucleótidos. Se identificaron niveles de corriente estadísticamente significativos dentro de estos eventos de translocación mediante el uso de una prueba t de Student, que se conoce comúnmente en la técnica por determinar la significación estadística entre valores vecinos (para más detalles, véase Carter et al., citado en este documento en otra parte, o John E. Freund, Mathematical Statistics, 5ª edición, Prentice Hall). Para las corrientes observadas de esta secuencia en particular, hubo casi el doble de niveles de corriente estadísticamente significativos identificados, ya que se procesaron los nucleótidos, con una topología (picos y depresiones de los niveles de corriente) de casi el doble de niveles entre cada pico y entre cada pico a través, medido por observación directa, que para un motor molecular de una sola etapa.

Para reducir el error experimental, se realizaron análisis de datos de la duración de la translocación de polinucleótidos a través de nanoporos en regiones de gran resolución de translocación de nucleótidos. En la secuenciación de cadenas y, en particular, la secuenciación en nanoporos de polinucleótidos, una región abásica puede producir una relación señal/ruido relativamente alta debido a la diferencia significativa en el flujo de iones que se bloquea cuando se compara con la de las secuencias de polinucleótidos vecinas. Por esta razón, los niveles estadísticamente significativos en la

vecindad de una región abásica pueden ser más probables debido al procesamiento de nucleótidos a través del nanoporo que a un efecto ulterior "ruidoso". Por esta razón, se seleccionaron las duraciones de 27 niveles de corriente que rodean e incluyen un pico de corriente abásica para el análisis de datos basado en la duración del nivel de corriente.

- 5 Las figuras 4A y 4B muestran el efecto ejemplar de la concentración de ATP en el tiempo de permanencia de las etapas de translocación fraccional, de acuerdo con algunas realizaciones. En la Fig. 4A, los niveles de corriente que rodean a un pico de corriente abásica de una secuencia de polinucleótidos que se transloca a través de un nanoporo con la inclusión de la etapa de translocación fraccional se marcaron secuencialmente de 1 a 27, y se representaron las duraciones medias.
- 10 Los niveles de corriente fueron detectados por un algoritmo que utiliza una prueba t de Student para determinar la significancia estadística entre los valores de corriente vecinos (ver Carter et al., citados en otra parte del presente documento). Otras técnicas son capaces de esto, incluyendo el umbral de velocidad y la minimización de chi cuadrado, y todas son conocidas en la técnica para los cambios de corriente relacionados con el procesamiento de nucleótidos, así como para la detección de etapas en el procesamiento de imágenes. Hubo una duración asociada con cada uno de estos niveles, y al comparar el mismo nivel sobre múltiples polinucleótidos de la misma secuencia que atraviesa el poro, se calculó la duración media de cada nivel. Estas duraciones medias fueron, por lo tanto, representativas de la duración típica asociada con cada nivel. Sin embargo, debido a la distribución exponencial de los tiempos de permanencia, las constantes de tiempo de estos tiempos de permanencia son más indicativas de su dependencia de ATP. Por este motivo, en la fig. 4B, los histogramas de duración de los niveles pares e impares (donde "par" e "impar" están asociados con los Índices de Nivel en la Figura 4A) se ajustaron a las curvas de decaimiento exponencial ($a \cdot e^{-t/\tau}$) y fueron representadas las constantes de tiempo para cada uno. Los histogramas de las duraciones de los niveles se construyeron incorporando la duración de cada nivel en cada evento de translocación de polinucleótidos en recipientes de tamaño equivalente. Luego, estos histogramas se ajustaron utilizando un algoritmo de ajuste de curvas disponible comercialmente (Matlab Curve Fitting Toolbox), que utiliza el método de mínimos cuadrados para ajustar los datos al modelo de decaimiento exponencial. Este método minimiza la suma de los residuos al cuadrado, donde un residuo se define como la diferencia entre un punto de datos y la respuesta ajustada a ese punto. Esta es una técnica estándar para ajustar datos a un modelo paramétrico. Las barras de error en la fig. 4B corresponden a los límites de confianza del 95% de cada ajuste.

- Como la fig. 4B muestra, el tiempo de permanencia para los niveles pares aumentó al disminuir la concentración de ATP, mientras que el tiempo de permanencia para los niveles impares se mantuvo constante. En consecuencia, el tiempo de permanencia para los niveles pares, que correspondió con las primeras etapas de translocación fraccional, se asoció ostensiblemente con la unión de ATP e inversamente proporcional a la concentración de ATP con una distribución exponencial, mientras que el tiempo de permanencia para los niveles impares, que correspondió con las segundas etapas de translocación fraccional, se asociaron ostensiblemente con la hidrólisis de ATP y dependencia del ATP.

Ejemplo III

Utilidad de las etapas de translocación fraccional en la secuenciación de polinucleótidos

- El Ejemplo III describe el aumento de la precisión de la secuencia mediante el uso de señales eléctricas obtenidas a partir de dos etapas de translocación fraccionaria de un ciclo de translocación completo en comparación con el uso de una única señal eléctrica obtenida a partir de un ciclo de translocación completo.

- Debido a que la "cabeza de lectura" de MspA es sensible a un tramo de 4 nucleótidos (4-mer) dentro de la zona de constricción, las trazas de corriente se generaron a partir de un mapa de cuadrómero que mide las corrientes correspondientes a todas las combinaciones de 4-mer observadas en el nanoporo de MspA. Para obtener más detalles sobre la medición de las corrientes correspondientes a las combinaciones de 4 mer, véase Laszlo et al., "Decoding long nanopore sequencing reads of natural DNA", Nature Biotechnology 32: 829-833 (2014). Sin embargo, debe tenerse en cuenta que diferentes poros pueden ser sensibles a diferentes números de nucleótidos dentro de las zonas de constricción. En el presente ejemplo, la precisión de la secuenciación se determinó comparando los resultados del Modelo de Markov Oculto (HMM) con la secuencia original de Bruijn como se describe a continuación. Para niveles de ruido experimentales típicos, tal como se ilustra en la FIG. 5 (es decir, ~0,5-2 pA, o de aproximadamente 0,5 a 1,5 pA), la precisión de la reconstrucción con etapas completas (diamantes) se redujo en comparación con las etapas fraccionales (cuadrados).

- Brevemente, los poros se establecieron con los métodos descritos anteriormente (ver Butler et al, Proc. Natl Acad Sci. USA, 105: 20647-20652 (2008); Manrao et al., PLoS ONE, 6: e25723 (2011)). Brevemente, se formaron bicapas lipídicas a través de una abertura horizontal de ~20 micras de diámetro en teflón a partir de 1,2-difitoil-sn-glicerol-3-fosfocolina (Avanti Polar Lipids). Los compartimentos en ambos lados de la bicapa contenían un tampón experimental de Hepes 10 mM, pH 8,0, KCl 400 mM, DTT 1 mM y MgCl₂ 10 mM. Se utilizó un Axopatch-200 B (Axon Instruments) para aplicar un voltaje a través de la bicapa (140 mV o 180 mV) y medir la corriente iónica. Se agregó MspA al compartimento cis conectado a tierra a una concentración de ~2,5 ng/ml. Una vez que se insertó una única proteína MspA en la abertura de teflón, el compartimento cis se lavó con tampón experimental para inhibir o evitar inserciones

adicionales. Todos los experimentos se realizaron a 23°C. La señal analógica de la corriente iónica se filtró en paso bajo a 20 kHz con un filtro Bessel de 4 polos y se digitalizó a 100 kHz utilizando un digitalizador 6363 de National Instruments. La adquisición de datos se controló con un software personalizado escrito en LabWindows/CVI (National Instruments). Los datos se analizaron con un software personalizado escrito en Matlab (The Mathworks). El ATP se usó típicamente a 1 mM, excepto para los experimentos de titulación de ATP, en cuyo caso las concentraciones de ATP oscilaron entre 10 μ M y 1 mM. El polinucleótido de translocación hibridado a un polinucleótido que contiene colesterol se usó a 10 nM. Se utilizó helicasa Hel308 Tga a una concentración final de 115 nM. Se agregaron polinucleótido y ATP a la cámara cis, seguido finalmente por helicasa Hel308 Tga. Alternativamente, puede emplearse un sistema de regeneración de ATP conocido en la técnica. El sistema ejemplar incluye ATP 2 mM, sal disódica de fosfato de creatina 10 mM, creatina quinasa 3,5 U/mL y pirofosfatasa inorgánica 0,6 U/mL.

La Fig. 5 representa la precisión de la reconstrucción de la secuenciación (Modelo de Markov oculto (HMM)) para etapas completas (diamantes) y medias etapas (cuadrados) trazas de corriente generadas in silico (descritas a continuación) con varios niveles de ruido agregado, de acuerdo con algunas realizaciones. La Fig. 5 muestra la precisión de la reconstrucción de la secuencia derivada de un análisis de algoritmo HMM/Viterbi de las trazas de bloqueo de corriente del modelo para una secuencia de Bruijn (256-mer). El algoritmo HMM general es, en algunos aspectos, similar a lo que se describe en Timp et al., *Biophys J.* 2012, 16 de mayo; 102 (10): L37-9. doi: 10.1016/j.bpj.2012.04.009. Este algoritmo puede recuperar el conjunto subyacente de 'estados' M de una serie de mediciones observadas. La forma básica de este algoritmo se basa en dos conjuntos de probabilidades determinadas experimentalmente: una probabilidad de "transición" de estado a estado y una probabilidad de "emisión" de observación de estado. Las medidas dadas en las etapas $i = 1, 2, 3, \dots, N$ para N mediciones. Un conjunto de probabilidades es la matriz de transición que describe la probabilidad de que durante un tiempo dado i , y un estado S_i (donde S es un estado en el conjunto de estados M), el estado posterior S_{i+1} (donde S_{i+1} no es necesariamente S_i). Para el sistema de nanoporos, con un nanoporo que es sensible a 4 nt, y está examinando los 4 nucleótidos canónicos (A, C, G, T), se obtienen $4^4 = 256$ estados, correspondientes a cada combinación de 4 nt. Cada uno de estos estados solo puede hacer la transición a uno de los 4 estados adyacentes.

La Fig. 6A representa transiciones de estado con probabilidad no nula necesaria para que un HMM decodifique la secuencia en un nanoporo en el que el polinucleótido es trasladado por una enzima motora, de acuerdo con algunas realizaciones. El motor es phi29 DNAP o un polinucleótido de traslación de enzima similar en etapas de 1 nucleótido. La Fig. 6B representa las transiciones de estado con una probabilidad no nula necesaria para que un HMM decodifique la secuencia en un nanoporo en el que el polinucleótido es trasladado por una enzima motora, de acuerdo con algunas realizaciones. El motor es la helicasa Hel308 o una enzima similar que permite el movimiento fraccional del polímero.

Las probabilidades de transición no cero para la matriz de transición de este sistema se muestran en la FIG. 6A para una enzima que se mueve en las etapas de un solo nucleótido. Usando este tipo de enzima, cada estado polinucleotídico o n-mer debe ir a uno de los 4 estados n-mer adyacentes. Para una enzima que toma una etapa de translocación fraccional, habrá más estados. A este respecto, un estado dado de etapa completo debe pasar a un estado de medio etapa (o etapa fraccional) antes de que se pueda observar otro estado de etapa completo. Por lo tanto, hay más estados disponibles con rutas más discernibles, lo que ayuda a la precisión de la caracterización de polinucleótidos.

El número de estados viene dado por $q \cdot 4^{n+1}$, donde n es el tamaño de lectura del nanoporo, y q es el número de etapas necesarias para completar un ciclo de translocación completo. Para $q = 2$ y $n = 4$, como se ve con la helicasa Hel308 y M2-NNN MspA, hay 2048 estados. La matriz de probabilidad de transición se describe gráficamente en la FIG. 6B para una enzima que se mueve en etapas de nucleótidos fraccionales. Cada estado correspondiente a un estado completo, puede hacer la transición a solo uno de los 1024 'medios estados' o 'estados fraccionados', mientras que cada uno de los estados medios (o fraccionales) puede hacer la transición a 4 estados diferentes, correspondientes a un nuevo estado dentro del encabezado de lectura del nanoporo. Para el algoritmo de decodificación HMM, se usa otro conjunto de probabilidad: la probabilidad de que una medición de corriente en el tiempo t , C_t , pertenezca al estado S_i . Este conjunto de probabilidades se determina experimentalmente o se estima a partir de observaciones experimentales anteriores. Dicha estimación se puede lograr mediante la aplicación iterativa de algoritmos de alineación, tal como se describe en Laszlo et al. 2014 (citado en otra parte del presente documento) o con la expectativa de maximizar un HMM. Para evaluar la utilidad de las etapas de translocación fraccional, se comparó la precisión de la secuenciación para una enzima con etapas de nucleótidos fraccionales con la de una enzima con etapas de un solo nucleótido. Se implementaron algoritmos de decodificación HMM Viterbi con software personalizado implementado en MATLAB, y se generaron 10 simulaciones de Monte-Carlo en experimentos in silico para cada condición. Las medias de secuenciación y las desviaciones estándar se obtuvieron a partir de las desviaciones promedio y estándar de estas 100 simulaciones de Monte-Carlo. Los niveles de corriente se generaron en base a los resultados de Manrao 2012 (citados en otra parte del presente documento). Se agregó ruido gaussiano, con anchos gaussianos dados por los valores mostrados en el eje X de la FIG. 5, para cambiar los valores de corriente observados in silico que se usaron en la reconstrucción de secuencias, y un experimento típico de secuenciación de nanoporos tiene una fluctuación de aproximadamente 1 pA_{rms} en los niveles promedio. Para un cambio Gaussiano agregado con un ancho de 0,5 pA, tanto la reconstrucción fraccional como la reconstrucción completa dieron lugar a una precisión de secuencia proporcional al 100%. Por encima de los cambios gaussianos agregados con anchos por encima de 0,5 pA, la precisión de la secuenciación para las etapas de translocación fraccional fue mayor que la precisión de la secuenciación para las etapas de la translocación no fraccional. Por lo tanto, la información adicional

de las etapas de translocación fraccional proporcionó o confirió una precisión de reconstrucción de secuencia mejorada cuando se agregaron más de 0,5 pA de ruido gaussiano a los niveles de corriente promedio.

Además de las fluctuaciones de ruido, los niveles omitidos causados por el movimiento estocástico de las enzimas reducirán la precisión de la secuenciación, o puede esperarse que lo hagan. Esta reducción en la precisión es, o puede ser, parcialmente compensada por la relectura de los patrones de nucleótidos en los cuadrantes adyacentes. Con la etapa de la translocación fraccional agregada hay una lectura adicional de los patrones de nucleótidos. Por ejemplo, la información sobre un k-mer dado se incluye en las etapas fraccionales vecinas, por lo que el k-mer se "re-lee" durante esas etapas fraccionarias vecinas. Por ejemplo, supongamos que un polinucleótido que tiene la secuencia ATCGTC está siendo translocado fraccionalmente a través de un nanoporo con una cabeza lectora sensible a 4 nucleótidos. Sin desear estar limitado por ninguna teoría, para un motor en etapas completas, el TCGT de 4-mer solo se lee cuando la región entre C y G está centrada en la cabeza lectora (es decir, la 'etapa' anterior tendría 'TC' centrado en la cabeza lectora, y solo se leería ATCG; la siguiente etapa tendría 'GT' centrado en la cabeza lectora y solo leería CGTC). Por lo tanto, si el motor omite la etapa de lectura TCGT, nunca se medirá ninguna información asociada con ese 4-mer en particular. Sin embargo, sin querer limitarse a ninguna teoría, con un motor de etapas fraccionales, como una helicasa Hel308, durante las etapas completas la región entre dos nucleótidos vecinos puede centrarse en la cabeza de lectura, mientras que durante las etapas fraccionales, los nucleótidos individuales pueden centrarse en la cabeza lectora. Entonces, cuando el 'CG' del polinucleótido mencionado anteriormente está centrado en la cabeza de lectura, se lee TCGT, tal como estaba en el caso de las etapas completas. La etapa fraccional anterior solo puede tener una C centrada en la cabeza lectora, y se puede leer la información sobre ATCGT; la siguiente etapa fraccional solo tiene G centrado en la cabeza lectora y se lee la información sobre TCGTC. Debido a que la información sobre 'TCGT' se puede leer 3 veces en el caso de etapas fraccionales y solo una vez en el caso de etapas completas, esta 're-lectura' adicional del 4-mer puede permitir obtener información sobre TCGT incluso si el motor omite una etapa asociada con él, lo que probablemente no es cierto en el caso de etapas completas. Para todas las fracciones de niveles eliminados, hay una mejora en la precisión de la secuenciación del 2-8%. Esto se demostró con simulaciones de Monte Carlo in silico adicionales que realizaban la eliminación aleatoria de los niveles de corriente. En conclusión, hubo un fuerte aumento en la precisión de la secuenciación para los modos de error vistos en los experimentos de secuenciación de nanoporos. Las figuras 10, 11 y 12 representan esquemas mediante los cuales se puede usar la información de etapas de translocación fraccional adicional para mejorar la precisión de la secuencia. Estos esquemas se ilustran adicionalmente a continuación en el Ejemplo VI. Los esquemas descritos son usos ejemplares y no pretenden ser limitativos.

Ejemplo IV

Utilidad de las etapas de translocación fraccional en la coincidencia de patrones

El Ejemplo IV describe el uso ejemplar de etapas de translocación fraccional para identificar niveles utilizando algoritmos conocidos. Usando un algoritmo de programación dinámico, como la alineación de Needleman-Wunsch, los niveles adicionales brindaron asistencia para encontrar patrones con precisión dentro de muchos niveles. Para obtener más detalles sobre los algoritmos de alineación de Needleman Wunsch, véase Durbin et al., Biological Sequence Analysis, ed. 11 (Cambridge University Press, Cambridge, Reino Unido 2006). Además de, o como alternativa al promedio de las corrientes de nivel, el uso de las duraciones de nivel, las desviaciones estándar de nivel de corriente, o las distribuciones de nivel, pueden mejorar aún más la precisión de coincidencia de patrones. En este ejemplo, se utilizó un algoritmo de alineación Needleman Wunsch para identificar los niveles correspondientes a una secuencia de 15 bases incrustada dentro de los niveles correspondientes a una secuencia de 1000 bases. Se comparó el uso de lo siguiente: (1) niveles correspondientes a movimientos de nucleótidos completos, (2) niveles correspondientes a movimientos de 2 medias etapas (o fraccionales), (3) niveles y duraciones correspondientes a movimientos de 2 medias etapas (o etapa fraccional). Se generaron observaciones in silico con 10 simulaciones de Monte Carlo con niveles cambiados por valores dados por valores aleatorios generados a partir de una distribución gaussiana de anchos variables. Los resultados se muestran en la fig. 7, que representa la precisión esperada de encontrar patrones de corriente en función del desplazamiento gaussiano, de acuerdo con algunas realizaciones. La desviación media y estándar de la precisión de alineación se generó a partir de la desviación media y estándar de las 10 simulaciones de Monte Carlo. En la Fig. 7, los diamantes representan un motor que tiene una etapa completa de nucleótidos, y los círculos representan un motor que tiene una etapa de translocación fraccional, y los cuadrados representan un motor que tiene una etapa de translocación fraccional combinado con valores de duración. En resumen, los patrones de nivel correspondientes a 15 nucleótidos se integraron dentro de los patrones de nivel correspondientes a una secuencia aleatoria de 1000 nucleótidos. Los niveles correspondían a un motor con etapas de nucleótidos completos (diamantes) como phi29 DNAP (solo etapa de translocación completa), o a un motor con una etapa de translocación fraccional (círculos), como Helicasa Hel308. Se usó la duración además de los valores de corriente para mejorar aún más las coincidencias (cuadrados). De los resultados en la fig. 7, se puede entender que para aumentar el ruido, la calidad coincidente fue considerablemente mayor para los algoritmos que utilizaban el movimiento de etapa de translocación fraccional. La calidad coincidente se mejoró aún más si también se utilizaban los valores de duración. Para comparar los niveles con el algoritmo de Needleman Wunsch, se utilizó una medida de similitud de nivel de entrada, o puntuación, para comparar niveles. En las pruebas se empleó una prueba t de Student para comparar los niveles de corriente. Para comparar (puntuar) la similitud de dos duraciones, se determinó la diferencia del logaritmo natural de las duraciones y se agregó a la puntuación otorgada por la prueba t de Student. El término "puntuación" se puede definir en la nomenclatura del algoritmo de Needleman Wunsch. Estas funciones de

puntuación representan ejemplos no limitativos de métodos que se pueden usar para comparar niveles de señal (por ejemplo, valores de corriente) y duraciones.

Ejemplo V

Modular las etapas fraccionales de helicasa

- 5 El Ejemplo V ejemplifica el uso de diferentes componentes de reacción para variar el tiempo de permanencia de la helicasa Hel308.

La Fig. 8 muestra la modulación ejemplar de la actividad helicasa Hel308 con concentraciones variables de pirofosfato, de acuerdo con algunas realizaciones. La Fig. 9 muestra la modulación ejemplar de la actividad helicasa Hel308 con el inhibidor de nucleótidos ortovanadato de sodio y con el análogo de nucleótido hidrato de la sal de adenosina 5'-(β,γ -imido) trifosfato de litio, de acuerdo con algunas realizaciones.

La actividad de helicasa Hel308 se moduló aumentando la concentración de pirofosfato. Brevemente, las condiciones de reacción fueron las descritas en el Ejemplo III con la inclusión de varias concentraciones de pirofosfato que variaban de 0 a 50 mM, por ejemplo, 0 mM (control), 5 mM, 10 mM, 20 mM, 30 mM, 40 mM y 50 mM. Los resultados se muestran en la fig. 8 e indican el porcentaje de actividad de helicasa en comparación con la actividad de helicasa en ausencia de pirofosfato (control). Las concentraciones de pirofosfato 5 y 10 mM dieron como resultado una disminución de la actividad de la helicasa a más del 75% del control. Las concentraciones de pirofosfato superiores a 10 mM dieron como resultado mayores disminuciones en la actividad de la helicasa y, por lo tanto, en el tiempo de permanencia de la helicasa. Se utilizó un ensayo de fluorescencia para controlar la capacidad de la helicasa para desenrollar el ADN dúplex. Un polinucleótido FRET de 49 nt (concentración final 50 nM) incluyó un grupo de 5' fluoresceína (/FAM/). Un polinucleótido que contenía un inactivador de 40 nt (50 nM final) incluyó un inactivador fluorescente, el inactivador de agujero negro (/BHQ1/). Estos dos polinucleótidos se hibridaron juntos calentándolos más allá de sus temperaturas de fusión a 75°C y enfriando lentamente a temperatura ambiente usando métodos conocidos en la técnica. El dúplex incluía un saliente 3' de 9 bases que se pudo unir a la helicasa de 3' a 5'. Un polinucleótido FRET de 40 nt complementario que era complementario al 100% del polinucleótido que contenía el inactivador de 40 nt estaba presente en un mínimo de 10 veces el exceso molar. Debido a que el inactivador y el fluoróforo estaban inicialmente muy cerca, la fluorescencia se inactivó. Sobre la base de que la helicasa desenrollaba el ADN dúplex, el polinucleótido que contenía el inhibidor de 40 nt tenía más probabilidades de unirse al polinucleótido de 40 nt de FRET complementario que de volver a unirse al polinucleótido de FRET de 49 nt. El nuevo polinucleótido FRET de 49 nt monocatenario, por lo tanto, emitió fluorescencia en presencia de una fuente de luz excitante adecuada. El tampón de ensayo incluía HEPES 10 mM, pH 8,0, KCl 400 mM, MgCl₂ 1 mM, DTT 1 mM, ATP 1 mM. La reacción se dejó proceder durante 20 minutos a temperatura ambiente antes de leer la fluorescencia.

La actividad de helicasa Hel308 y, por lo tanto, el tiempo de permanencia también se mostró reducido en presencia de cualquiera de los inhibidores de nucleótidos o análogos de ortovanadato de sodio e hidrato de sal de adenosina 5'-(β,γ -imido)trifosfato de litio, respectivamente. Brevemente, las condiciones de reacción fueron las descritas en el Ejemplo III con la inclusión de ortovanadato de sodio ("NaOrtovanadato" en la Figura 9) o hidrato de sal de adenosina 5'-(β,γ -imido) trifosfato de litio ("AMP-PNP" en la Figura 9) a una concentración de 5 mM de inhibidor o análogo. Los resultados se muestran en la fig. 9 e indican el porcentaje de actividad de helicasa en comparación con la actividad de helicasa en ausencia del inhibidor de nucleótidos o análogo (control). Las concentraciones de inhibidor o análogo de 5 mM dieron como resulta una disminución de la actividad de helicasa a más del 85% del control y, por lo tanto, se puede esperar que aumenten el tiempo de permanencia de la helicasa, o el tiempo que tarda la helicasa en moverse a lo largo del ADN. Por ejemplo, un aumento en el tiempo de permanencia puede alargar el tiempo de una etapa fraccional, lo que permite más tiempo para adquirir una señal.

Ejemplo VI

Métodos para procesar la información de las etapas fraccionales para mejorar la precisión de la secuenciación

- 45 El Ejemplo VI ilustra tres métodos para procesar la información adicional obtenida de las etapas de translocación fraccional para mejorar la precisión de la secuenciación.

La Fig. 10 ilustra un método ejemplar para procesar la información adicional obtenida de las etapas de translocación fraccional utilizando la información de la corriente de nivel y la duración. El método se puede aplicar a dos lecturas de secuencia independientes. Usando este esquema, la traza de corriente se somete a un algoritmo de detección de pasos, en donde se encuentran los niveles de corriente y la duración de esos niveles. Basado, al menos en parte, en la duración de los niveles, un HMM de dos estados identifica los niveles como una etapa completa (larga) o como una etapa media (o fraccional) (corta), o como un salto potencial dentro de las observaciones. Estas etapas largas y cortas identificadas y la información de salto se usan luego por un HMM, Viterbi o un algoritmo de coincidencia de patrones, o una combinación adecuada de los mismos, para reconstruir la secuencia de polinucleótidos para los dos tipos de estados por separado (largo y corto, respectivamente, correspondiente a la etapa completa) y los niveles de etapa media (o fraccional)). Las secuencias resueltas luego se comparan y se usan para mejorar la precisión de la secuenciación de polinucleótidos, por ejemplo, ajustando el algoritmo HMM, Viterbi o de coincidencia de patrones. La alineación se puede usar para identificar ubicaciones que coinciden mal en las dos lecturas de secuencia

independientes.

La Fig. 11 ilustra un método ejemplar para procesar la información adicional obtenida de las etapas de translocación fraccional utilizando la información de nivel de corriente y la duración. El método se puede aplicar a dos lecturas de secuencia concurrentes. En este método, la traza de corriente se somete primero a un algoritmo de detección de etapas para encontrar los niveles. La media (o mediana) de los valores de corriente de nivel y la duración de cada nivel se ingresan, como un par, en un algoritmo HMM bidimensional, Viterbi o de coincidencia de patrones, o una combinación adecuada de los mismos, que examina la duración y los valores de corriente y estimaciones o resoluciones de secuencia óptima para estados medios (o fraccionales) y estados completos. En esta técnica, la probabilidad de emisión de HMM es bidimensional: $P_i(\text{emisión}_i) = P_i(\text{cur}_i, \text{dur}_i) = P_i(\text{cur}_i) * P_i(\text{dur}_i)$ donde i es un "estado" que corresponde a una etapa larga o etapa de translocación del polinucleótido corta (fraccional), y cur_i y dur_i son, respectivamente, el nivel de corriente y la duración para el número de nivel i . El HMM bidimensional puede tomar como entrada un mapa de consenso y una distribución de probabilidad para niveles largos (estados completos), y un mapa de consenso y una distribución de probabilidad para niveles cortos (estados medios (o fraccionales)). El HMM bidimensional puede proporcionar como resultado una resolución de la secuencia de nucleótidos.

La Fig. 12 ilustra un método ejemplar para procesar la información adicional obtenida de las etapas de translocación fraccional usando trazas de corriente directamente. El método puede aplicarse con o sin el uso de información de duración. En referencia al uso de la información de duración, en este método la traza de corriente se analiza directamente con un HMM dependiente de la duración. En esta versión de un HMM, la duración de un nivel se determina al mismo tiempo que la secuencia más probable y el estado de etapa completo o medio (o fraccional). Si el estado se ha mantenido sin cambios entre dos iteraciones de tiempo, la duración del estado dado aumentará. Esta duración se utiliza para mejorar la evaluación de si ese estado se encuentra en estado completo o fraccional.

Ejemplo VII

Métodos adicionales para procesar información de etapas fraccionales para mejorar la precisión de la secuenciación

El Ejemplo VII describe métodos ejemplares adicionales para procesar información de etapas fraccionales para mejorar la precisión de la secuenciación.

Los modelos ocultos de Markov (HMM) y los algoritmos de Viterbi se han utilizado anteriormente para la resolución de bases en función de las señales de los polinucleótidos que se desplazan a través de los nanoporos que demandan motores moleculares de un solo paso. Para obtener más detalles, véase Timp et al., "DNA Base-Calling from a Nanopore Using a Viterbi Algorithm", Biophysical Journal 102: L37-L39 (mayo de 2012). La Fig. 19A ilustra esquemáticamente un aspecto de un modelo oculto de Markov (HMM) utilizado para caracterizar una señal de translocación de un solo etapa de un polinucleótido a través de un poro, por ejemplo, en el que un nivel de señal dado corresponde a la translocación de un nucleótido a través de un poro, por ejemplo, por una polimerasa o una helicasa. Como se indica en otra parte del presente documento, un nivel de señal puede no corresponder necesariamente a la presencia de un solo nucleótido dentro de la constricción de un poro, sino que puede corresponder a la presencia de una "palabra" que incluya una pluralidad de nucleótidos, por ejemplo, dos, tres, cuatro, cinco, seis, siete, ocho, nueve, diez o más de diez nucleótidos. Dicha "palabra" también puede denominarse "k-mer". En la realización ilustrada en la fig. 19A, las "palabras" o "k-mers" tienen una longitud de cuatro nucleótidos o son "cuadrangulares" o "4-mers", correspondientes a un nivel de señal que se basa en la presencia de cuatro nucleótidos en la constricción de un poro.

En la Fig. 19A, se puede ver que para una posición dada i de un polinucleótido que se transloca a través de un poro, un cuadrómero dado en la constricción del poro puede incluir cualquier combinación posible de cuatro nucleótidos, por ejemplo, AAAA, AAAC, AAAG, AAAT, ... TTTT. Es posible que no sea necesariamente posible en función del nivel de señal correspondiente a dicho cuadrómero para identificar de forma única a ese cuadrómero. Por ejemplo, dos cuadrómeros diferentes, por ejemplo, dos cuadrómeros diferentes que sean adyacentes entre sí en la secuencia esperada, potencialmente pueden tener los mismos niveles de señal entre sí. Timp describe valores de corriente ejemplares para tripletes de ADN (3-mers), en base a los cuales se puede observar que ciertos tripletes pueden tener los mismos niveles de señal entre sí, lo que inhibe la resolución de la base de nucleótidos de las bases en ese triplete basándose únicamente en el nivel de corriente correspondiente a ese triplete. Debe entenderse que ciertos cuadrómeros (y, más generalmente, ciertos k-mers) pueden tener niveles de señal que son indistinguibles entre sí, lo que inhibe la resolución de la base de nucleótidos de las bases en ese cuadrómero o k-mer basado únicamente en el nivel de corriente correspondiente a ese 4-mer o k-mer. En consecuencia, utilizando la terminología de los HMM, las bases en tales cuadrómero o k-mers que son indistinguibles entre sí en función de la observación del nivel de la señal pueden modelarse como un "estado oculto".

Se puede usar información adicional basada en observaciones de otras posiciones de un solo etapa del polinucleótido en la constricción de poros para aumentar la probabilidad de identificar las bases en ese cuadrómero o k-mer, y así identificar con precisión el "estado oculto". Por ejemplo, en la fig. 19A, también se puede ver que, para la siguiente posición $i + 1$ de un polinucleótido que se transloca a través de un poro, un cuadrómero dado en la constricción del poro puede tener solo ciertas combinaciones posibles de cuatro nucleótidos porque los últimos tres nucleótidos de la posición i corresponden a los tres primeros nucleótidos de la posición $i + 1$. Como tales, las mediciones de las señales para los estados i e $i + 1$ se pueden usar para aumentar la probabilidad de identificar

correctamente los cuadrómeros presentes en una o ambas posiciones i y $i + 1$ (o, de manera equivalente, las posiciones $i - 1$ y i). Por ejemplo, según la secuencia AAAA correspondiente a la posición i del polinucleótido, solo las cuatro secuencias AAAA, AAAC, AAAG y AAAT están disponibles para la posición $i + 1$. Las cuatro secuencias disponibles en la posición $i + 1$ pueden identificarse fácilmente para cada secuencia posible en la posición i . Análogamente, en función de la secuencia en la posición $i + 1$ del polinucleótido, las cuatro secuencias disponibles en la posición $i + 2$ del polinucleótido pueden identificarse fácilmente. Un algoritmo de Viterbi para un motor de un solo paso, en el que hay una correspondencia uno a uno entre los niveles de señal y las posiciones $i, i + 1, i + 2, \dots, i + n$, donde n es el número de nucleótidos en un polinucleótido, puede expresar la señal del conjunto de niveles ordenados $L = \{l_1, l_2, \dots, l_n\}$. Cada nivel l_i , que corresponde a la posición i del polinucleótido, puede expresarse como uno o más de la media de ese nivel de señal ($media_i$), la desviación estándar de ese nivel de señal (std_i), o la duración de ese nivel de señal (dur_i). El conjunto de posibles cuadrómeros se puede expresar como $prev(q) = \{q_1, q_2, \dots, q_4\}$ que define los valores posibles del cuadrómero correspondiente a la ubicación anterior del polinucleótido (la posición $i - 1$), cuando el cuadrómero correspondiente a la ubicación presente del polinucleótido (la posición i) es q . Por ejemplo, $prev(AACC) = \{AAAC, CAAC, GAAC, TAAC\}$. Basándose en el nivel de señal observado Q_i correspondiente al nivel observado jth , la puntuación de probabilidad para un cuadrómero q dado que está presente en la posición i se puede expresar como:

$$Puntuación(nivel, i, q) = \max \left\{ \begin{array}{l} Puntuación(nivel, i-1, q) + InsPen \\ \max_{p \in prev(q)} puntuación(nivel, i-1, q) + s(l_i | q) \\ \max_{p \in prev(q)} puntuación(nivel, 1, p) + DelPen \end{array} \right\} \quad (1)$$

donde $s(l_i | q)$ corresponde a una compensación que representa la probabilidad de observar el nivel l_i dado el cuadrómero q , $InsPen$ es una penalización de inserción (una penalización que corresponde a un nivel de señal que se observa pero no corresponde a un cuadrómero en el polinucleótido), y $DelPen$ es una penalización por eliminación (una penalización correspondiente a un cuadrómero en el polinucleótido pero no tiene un nivel de señal correspondiente).

La Fig. 19B ilustra esquemáticamente un aspecto de un HMM ejemplar utilizado para caracterizar señales de la translocación en etapas fraccionales de un polinucleótido a través de un poro utilizando una helicasa Hel308, de acuerdo con algunas realizaciones. En la Fig. 19B, nuevamente se puede ver que para una posición dada i de un polinucleótido que se transloca a través de un poro, un cuadrómero dado en la constricción del poro puede incluir cualquier combinación posible de cuatro nucleótidos, por ejemplo, AAAA, AAAC, AAAG, AAAT, ...TTTT. Se puede usar información adicional basada en observaciones de etapas fraccionales de posiciones de etapas fraccionales, así como otras posiciones de un solo paso, del polinucleótido en la constricción de poros para aumentar la probabilidad de identificar con precisión las bases en ese cuadrómero o k -mer, y, por lo tanto, de identificar con precisión el "estado oculto", con una precisión mejorada en relación con el uso de solo posiciones de un solo paso.

Por ejemplo, en la fig. 19B, también se puede ver que, para un motor etapa a etapa fraccional, la siguiente posición del polinucleótido que se transloca a través de un poro es "i fraccional", y en la cual un cuadrómero dado en la constricción del poro puede tener solo ciertas combinaciones posibles de cuatro nucleótidos, porque los últimos tres nucleótidos de la posición i corresponden a los tres primeros nucleótidos de la posición "i fraccional". Como tal, las mediciones de las señales para los estados fraccionales i e $i + 1$ se pueden usar para aumentar la probabilidad de identificar correctamente los cuadrómeros presentes. Por ejemplo, en función de la secuencia AAAA correspondiente a la posición i del polinucleótido, solo las cuatro secuencias AAAA, AAAC, AAAG y AAAT están disponibles para la posición i fraccional. Las cuatro secuencias disponibles en la posición i se pueden identificar fácilmente para cada secuencia posible en la posición i .

Adicionalmente, en la fig. 19B, también se puede ver que, para la siguiente posición $i + 1$ llena de un polinucleótido que se transloca a través de un poro, cuya posición sigue inmediatamente i fraccional, un cuadrómero dado en la constricción del poro puede tener solo una secuencia posible, porque los cuatro nucleótidos de la posición $i + 1$ completa corresponden a los mismos nucleótidos que para la posición i fraccional. Como tal, las mediciones de señales correspondientes a las posiciones i, i fraccional e $i + 1$ completa se pueden utilizar para aumentar la probabilidad de identificar correctamente los cuadrómeros presentes en algunas o todas las posiciones fraccionales i, i fraccional e $i + 1$ completa (o, las posiciones equivalente, $i - 1$ e i). Por ejemplo, según la secuencia AAAA correspondiente a la posición i del polinucleótido, solo las cuatro secuencias AAAA, AAAC, AAAG y AAAT están disponibles para la posición i fraccional y para $i + 1$ completa. Un algoritmo de Viterbi modificado para un motor de etapa fraccional, en el que existe una correspondencia entre los niveles de señal y las posiciones de etapa fraccional y de etapa completa i, i fraccional, $i + 1$ completo, $i + 1$ fraccional, $i + 2$ completo, $i + 2$ fraccional, ... $i + n$ fraccional, $i + n$ completo, donde n es el número de nucleótidos en un polinucleótido, puede expresar los niveles de señal l como el conjunto de niveles. Análogamente, como se discutió anteriormente con referencia a la FIG. 19A, cada nivel de señal l correspondiente a la posición i completa o i fraccional puede expresarse como uno o más de la media de ese nivel de señal ($media_i$), la desviación estándar de ese nivel de señal (std_i) o la duración de ese nivel de señal (dur_i). Dado el cuadrómero q en la etapa de translocación fraccional de corriente, el conjunto de posibles cuadrómeros correspondientes a la etapa de translocación completa anterior se puede definir como $prev(q) = \{q_1, q_2, \dots, q_4\}$. Por ejemplo, $prev(AACC) = \{AAAC, CAAC, GAAC, TAAC\}$.

Con base en el nivel de señal observado l_i correspondiente a la posición i , la puntuación de probabilidad $puntuación_f$ para un cuadrómero dado q , que está presente en la posición i para una ubicación correspondiente a un estado de translocación completa, y la puntuación de probabilidad $puntuación_h$ para un cuadrómero dado q , que está presente en la posición i para una ubicación correspondiente a un estado de translocación medio (o fraccional), se puede expresar de la siguiente manera:

$$Puntuación_f(nivel_i, i, q) = \max \left\{ \begin{array}{l} Puntuación_f(nivel, i-1, q) + InsPen \\ \max_{p \in prev(q)} puntuación_h(nivel, i-1, q) + s_f(l_i | q) \\ \max_{p \in prev(q)} puntuación_h(nivel, 1, p) + DelPen \end{array} \right\} \quad (2)$$

$$Puntuación_h(nivel_i, i, q) = \max \left\{ \begin{array}{l} Puntuación_h(nivel, i-1, q) + InsPen \\ \max_{p \in prev(q)} puntuación_f(nivel, i-1, p) + s_h(l_i | p, q) \\ \max_{p \in prev(q)} puntuación_f(nivel, 1, p) + DelPen \end{array} \right\} \quad (3)$$

donde $s_f(l_i | q)$ corresponde a una compensación que representa la probabilidad de observar el nivel l_i dado el cuadrómero q en el estado de translocación total, $s_h(l_i | p, q)$ corresponde a una compensación que representa la probabilidad de observar el nivel l_i dado el cuadrómero q y anterior cuadrómero p en estado translocacional fraccional, $InsPen$ es una penalización de inserción (una penalización que corresponde a un nivel de señal que se observa pero no corresponde a un cuadrómero en el polinucleótido), y $DelPen$ es una penalización por eliminación (una penalización que corresponde a un cuadrómero en el polinucleótido pero que no tiene un nivel de señal correspondiente).

Además, la programación dinámica se puede utilizar para la coincidencia de patrones para un motor molecular de etapas fraccionales (como una helicasa Hel308). La comparación dinámica de patrones se describe para un motor molecular de un solo etapa en Laszlo et al., "Decoding long nanopore sequencing reads of natural DNA", Nature Biotechnology 32: 829-833 (2014). Por ejemplo, para un motor molecular de un solo paso, los niveles de señal l pueden expresarse como el conjunto de niveles $L = \{l_1, l_2, \dots, l_n\}$, donde cada nivel de señal l_i correspondiente a la posición de la etapa de translocación completa del polinucleótido puede expresarse como uno o más de la media de ese nivel de señal ($media_i$), la desviación estándar de ese nivel de señal (std_i) o la duración de ese nivel de señal (dur_i). Basándose en el nivel de señal observado l_i , la puntuación de probabilidad para un cuadrómero q_i determinado que se mide puede expresarse como:

$$Puntuación(i, j) = \max \left\{ \begin{array}{l} Puntuación(i-1, j) + InsPen \\ Puntuación(i-1, j-1) + s(l_i | q_j) \\ Puntuación(i, j-1) + DelPen \end{array} \right\} \quad (4)$$

donde i representa la posición en la secuencia de nivel; j representa la posición en la secuencia de ADN, la última base de cuadrómero q_j sería la base en la posición j ; la puntuación (i, j) representa cómo de buena es la coincidencia entre los niveles $l_1, \dots, l_i$ y los cuadrómeros $q_1, \dots, q_j$; $s(l_i | q_j)$ corresponde a una compensación que representa la probabilidad de observar el nivel l_i dado el cuadrómero q_j ; $InsPen$ es una penalización de inserción (una penalización correspondiente a un nivel de señal que se observa pero no corresponde a un cuadrómero en el polinucleótido); $DelPen$ es una penalización por eliminación (una penalización correspondiente a un cuadrómero en el polinucleótido pero que no tiene un nivel de señal correspondiente).

Para un motor molecular de etapa fraccional como Hel308, los niveles de señal l pueden expresarse como el conjunto de niveles $L = \{l_1, l_2, \dots, l_n\}$, donde cada nivel de señal l_i correspondiente a la posición i completa o i fraccional del polinucleótido puede expresarse como uno o más de la media de ese nivel de señal ($media_i$), la desviación estándar de ese nivel de señal (std_i), o la duración de ese nivel de señal (dur_i). Basándose en el nivel de señal l_i observado, la puntuación de probabilidad $puntuación_f$ para un cuadrómero q_i dado que se mide correspondiente a un estado de translocación total, y la puntuación de probabilidad $puntuación_h$ para un cuadrómero q_i dado que se mide en un estado de translocación medio (o fraccional), puede expresarse de la siguiente manera:

$$Puntuación_f(i,j) = \max \left\{ \begin{array}{l} Puntuación_h(i-1, j) + InsPen \\ Puntuación_h(i-1, j-1) + s_f(l_i|q_j) \\ Puntuación_f(i, j-1) + DelPen \end{array} \right\} \quad (5)$$

$$Puntuación_h(i,j) = \max \left\{ \begin{array}{l} Puntuación_f(i-1, j) + InsPen \\ Puntuación_f(i-1, j-1) + s_h(l_i|q_{j-1}q_j) \\ Puntuación_h(i, j-1) + DelPen \end{array} \right\} \quad (6)$$

donde i representa la posición en la secuencia de nivel; j representa la posición en la secuencia de ADN, la última base de cuadrómero q_j sería la base en la posición j ; $puntuación_f(i,j)$ y $puntuación_h(i,j)$ representan cómo es la coincidencia entre los niveles $l_1, \dots, l_i$ y los cuadrómeros $q_1, \dots, q_j$, respectivamente, asumiendo un estado completo o fraccional; $s_f(l_i|q_j)$ y $s_h(l_i|q_j)$ corresponden a compensaciones que representan la probabilidad de observar el nivel l_i , dado el cuadrómero q_j en estado completo y fraccional, respectivamente; $InsPen$ es una penalización de inserción (una penalización correspondiente a un nivel de señal que se observa pero no corresponde a un cuadrómero en el polinucleótido); $DelPen$ es una penalización por eliminación (una penalización correspondiente a un cuadrómero en el polinucleótido pero que no tiene un nivel de señal correspondiente).

Algunos ejemplos de resultados de secuenciación de novo utilizando etapas fraccionales se describirán a continuación con referencia a la FIG. 20A. Se generó una biblioteca de polinucleótidos de 75-500 mer con base en el ADN humano, y los datos de nanoporos se recopilaron de manera análoga a como se describe en otras partes del presente documento con referencia a los Ejemplos II y III. La resolución de bases de nucleótidos basada en los datos se analizó utilizando el algoritmo Viterbi modificado descrito utilizando las ecuaciones (2) y (3) anteriores. La secuencia resuelta de bases luego se alineó con un conjunto de 150-500-mers, 75 de los cuales eran los verdaderos 500-mers y 75 de los cuales eran secuencias de 500-mer "señuelo" o "simulacro". En la Fig. 20A, que ilustra la longitud de lectura en función de la precisión de la alineación (utilizando un alineador LASTAL como se describe en Kielbasa et al., "Adaptive seeds tame genomic sequence comparison", Genome Research 21: 487-493 (2011)), los diamantes abiertos corresponden a resultados en los que la secuencia resuelta de bases se alinea con las secuencias correctas ("diana"), y los diamantes cerrados corresponden a resultados en los que la secuencia resuelta de bases se alinea a una secuencia "señuelo" o "simulacro". Se puede entender a partir de la fig. 20A que, para longitudes de lectura superiores a aproximadamente 200 pares de bases, se pueden obtener precisiones superiores a aproximadamente el 60%. La precisión adicional potencialmente puede aumentarse utilizando técnicas conocidas, como leer ambas cadenas del ADN.

Algunos ejemplos de resultados de coincidencia de patrones utilizando etapas fraccionales se describirán a continuación con referencia a las Figs. 20B-20C. La misma biblioteca de 75 polinucleótidos de 500 mer y el mismo protocolo experimental se usó como se describió anteriormente con referencia a la FIG. 20A. La resolución de bases de nucleótidos basada en los datos se analizó mediante programación dinámica para la comparación de patrones descrita utilizando las ecuaciones (5) y (6) anteriores. La secuencia resuelta de bases luego se alineó con un conjunto de 150-500-mers, 75 de los cuales eran los verdaderos 500-mers y 75 de los cuales eran secuencias de 500-mer "señuelo" o "simulacro". En la Fig. 20B, que ilustra el tamaño de la alineación en función de la puntuación de alineación, los diamantes abiertos corresponden a los resultados en los que la secuencia resuelta de bases se alinearon con las secuencias correctas ("diana"), y los diamantes cerrados corresponden a los resultados en los que la secuencia de bases resuelta fue alineada con una secuencia "señuelo" o "simulacro". Se puede entender a partir de la fig. 20B que se pueden obtener puntuaciones de alineación superiores a aproximadamente 40 para tamaños de alineación superiores a aproximadamente 200 pares de bases. En la Fig. 20C, que también ilustra el tamaño de la alineación en función de la puntuación de alineación, los diamantes abiertos corresponden a los resultados en los que la secuencia de bases resuelta se alineó con las secuencias correctas ("diana"), y los diamantes cerrados corresponden a los resultados en los que la secuencia de bases resuelta se alinea con una secuencia de "señuelo" o "simulacro". Se puede entender a partir de la fig. 20C que se pueden obtener puntuaciones de alineación mayores que aproximadamente 20 para tamaños de alineación mayores que aproximadamente 50 pares de bases. Se puede observar que el modelo de etapas fraccionales puede identificar con más precisión más eventos que el modelo de un solo paso.

Además, se observó que, para un evento de translocación con 1332 niveles, la comparación de patrones (Ecuaciones 5 y 6) contra un conjunto de datos de 80 kb llevó aproximadamente 145 segundos en un solo hilo, mientras que, para el mismo evento con 1332 niveles, la secuenciación de novo (ecuaciones 2 y 3) contra ese conjunto de datos llevó aproximadamente 69 segundos en un solo hilo. Se observó que la complejidad de la coincidencia de patrones crece linealmente con el conjunto de datos de nucleótidos, mientras que la complejidad de la secuenciación de novo fue independiente del conjunto de datos. Se observó una coincidencia de patrones para identificar con precisión los eventos más cortos, que la secuenciación de novo no pudo identificar. Además, se observó que el modelo de etapas fraccionales para la coincidencia de patrones produce más positivos verdaderos que un modelo de un solo paso, lo que indica que el modelo de etapas fraccionales puede ser un mejor modelo para explicar los datos de helicasa.

Ejemplo VIII

Etapas de translocación fraccional con helicasas Hel308 adicionales

El Ejemplo VIII describe las etapas de translocación fraccional observadas con helicasas Hel308 ejemplares utilizadas como motores moleculares.

- 5 Los experimentos para el Ejemplo VIII se realizaron de forma análoga a como se describió anteriormente con referencia al Ejemplo I, usando un único nanoporo MspA 2NNN en una bicapa lipídica DphPC, y usando los parámetros enumerados a continuación en la Tabla 3, donde "Hel308 Mbu (A)" se refiere a un conjunto de los parámetros utilizados para un primer experimento utilizando Hel308 Mbu y "Hel308 Mbu (B)" se refiere a un conjunto de parámetros utilizados para un segundo experimento utilizando Hel308 Mbu. Las bicapas lipídicas se formaron a partir de 1,2-difitoil-sn-glicero-3-fosfolina (lípidos polares de Avanti). La bicapa cubría una abertura horizontal de aprox. 20 micras de diámetro en teflón. Se agregó M2-NNN-MspA al lado conectado a tierra de la bicapa a una concentración de ~2,5 ng/ml. Una vez que se insertó un solo poro, el compartimiento se lavó con un tampón experimental para evitar nuevas inserciones. Un amplificador de pinza de parche Axopatch-200B (Axon Instruments) aplicó un voltaje a través de la bicapa de 180 mV y midió las corrientes iónicas. La señal analógica se filtró a paso bajo a 50 kHz con un filtro Bessel de 4 polos y luego se digitalizó a cinco veces la frecuencia del filtro a paso bajo. La adquisición de datos se controló con software personalizado escrito en LabWindows/CVI (National Instruments). Los compartimentos de ~60 µl a ambos lados de la bicapa contenían un tampón experimental de la concentración apropiada de KCl, EDTA 1 mM, DTT 1 mM, ATP 1 mM, MgCl₂ 5 mM y HEPES/KOH 10 mM tamponados a pH 8,0. Se utilizó helicasa natural Mbu Hel 308 a la concentración indicada como motor molecular.
- 10
- 15
- 20 En ambos experimentos de Hel308 Mbu y el experimento de Hel308 Tga, el ADN se leyó en la dirección de 3' a 5', mientras que en el experimento phi29 polimerasa, el ADN se leyó en la dirección de 5' a 3'.

Tabla 3

Parámetro	Hel308 Mbu (A)	Hel308 Mbu (B)	Hel308 Tga	phi29
[KCl]	300 mM	600 mM	400 mM	300 Mm
pH	8,0	8,0	8,0	8,0
[MgCl ₂]	5 mM	5 mM	5 mM	5 mM
[ATP]	1mM	1mM	1 mM	-
[DTT]	1 mM	1mM	1mM	1mM
[EDTA]	1mM	1mM	1mM	1mM
[DNA]	10 nM	10 nM	10 nM	10 nM
[motor molecular]	150 nM	150 nM	150 nM	1,5 µM
Voltaje	180 mV	180 mV	180 mV	180 mV

- 25 Las figuras 17A-17D muestran la comparación de los eventos de translocación de la helicasa Hel308 Mbu, helicasa Hel308 Tga y polimerasa Phi29 utilizando ciertos parámetros, según algunas realizaciones. La Fig. 17A muestra las etapas de translocación observadas con la helicasa Hel308 Mbu utilizando los parámetros "Hel308 Mbu (A)" mostrados en la Tabla 3. El polinucleótido de translocación (SEC ID NO: 72:/5Phos/AAACCTTCCXCCCGTACCGTGCCGTACCGTTCCGTACCGTATTTTTTCTCACTATCGCATTCTCATGCAGGTCGTAGCC, donde X = abásico) se hibridó a un polinucleótido que contenía colesterol (SEC ID NO: 73: AAAAAAATACGGTACGGAACGGAACGGTACGGCACGGTACGGGTTTTTTTTTTTTTTT/3CholTEG). La Fig. 17B muestra las etapas de translocación observadas con la helicasa Hel308 Mbu usando los parámetros "Hel308 Mbu (B)" mostrados en la Tabla 3 y usando las mismas secuencias de polinucleótidos que en la FIG. 17A. La Fig. 17C muestra las etapas de translocación observadas con una helicasa Hel308 Tga usando los parámetros de "Hel308 Tga" mostrados en la Tabla 3 y usando las mismas secuencias de polinucleótidos que en la FIG. 17A. La Fig. 17D muestra las etapas de translocación observadas con la polimerasa phi29 utilizando los parámetros "phi29" mostrados en la Tabla 3 y utilizando las mismas secuencias de polinucleótidos que en la FIG. 17A; la gráfica phi29 en la fig. 17D se reflejó sobre el eje vertical para facilitar las comparaciones entre las FIGS. 17A, 17B, 17C y 17D.
- 30
- 35

Puede verse en las figs. 17A-17D que, para la secuenciación con cada helicasa, el nanoporo detectó las características generalmente designadas como "a" (que corresponden a un valle en la señal) y "b" (que corresponden a un pico en la

señal). También se puede ver que, para la secuenciación con la helicasa Hel308 Tga Hel308 (Figura 17C), se observó aproximadamente el doble de niveles que para la helicasa phi29 (Figura 17D). También se puede ver que, para la secuenciación con la helicasa Hel308 Mbu en las condiciones "Hel308 Mbu (B)" (FIG. 17B), se observó un mayor número de niveles que para la helicasa Hel308 Mbu bajo las condiciones "Hel308 Mbu (A)" (Figura 17a). También se puede ver que, para la secuenciación con la helicasa Hel308 Mbu en las condiciones "Hel308 Mbu (B)" (Figura 17B), se observaron menos niveles que para la helicasa Hel308 Tga (Figura 17C), pero se observaron más niveles en cuanto a la helicasa phi29 (figura 17D). Las figuras 17A-17D puede interpretarse como que significa (1) que múltiples variantes de la helicasa Hel308 (por ejemplo, tanto Tga como Mbu) muestran una etapa fraccional, mientras que no se observó ninguna etapa fraccional para la polimerasa Phi29; y (2) que las etapas fraccionales se pueden dilucidar cambiando una variable o parámetro ambiental, por ejemplo, la concentración de KCl. Además, otros datos indican que la duración de los niveles aumenta con una disminución en la concentración de ATP cuando se utiliza Mbu, por ejemplo, que las duraciones y, por lo tanto, los mecanismos físicos de las etapas fraccionales en Mbu también pueden ser dependientes de ATP.

Ejemplo IX

Uso de estresores, opcionalmente en combinación con multimodalidad

Como debe quedar claro en base a la descripción proporcionada en este documento, muchas variables o parámetros ambientales pueden afectar a la forma en que un sistema de nanoporos lee o genera una señal basada en una secuencia de polinucleótidos en particular. Las variables o parámetros ejemplares que pueden proporcionar tal efecto pueden incluir la temperatura, la concentración de sales (p. Ej., Mg, Cl), la concentración del cofactor (p. ej., ATP), la concentración de productos de ATP como pirofosfato, pH, el motor molecular particular utilizado (p. ej., la helicasa Hel308 particular utilizada), presión y similares.

Por ejemplo, tal como se describió anteriormente con referencia al Ejemplo II y las Figs. 4A y 4B, la concentración de ATP puede afectar el tiempo de permanencia de los niveles correspondientes a ciertas etapas de translocación. Por ejemplo, se observó que los tiempos de permanencia para las primeras etapas de translocación fraccional aumentaron con la disminución de la concentración de ATP, y se asocian ostensiblemente con la unión de ATP y son inversamente proporcional a la concentración de ATP. Como otro ejemplo, tal como se describió anteriormente con referencia al Ejemplo V y la FIG. 8, la concentración de pirofosfato puede afectar a la actividad de la helicasa Hel308. Por ejemplo, se observó que la actividad de la helicasa Hel308 disminuía al aumentar la concentración de pirofosfato, aumentando así el tiempo de permanencia de la helicasa. Como otro ejemplo, tal como se describió anteriormente con referencia al Ejemplo V y la FIG. 9, la concentración del inhibidor de nucleótidos o análogo puede afectar a la actividad de la helicasa Hel308. Por ejemplo, se observó que la actividad de la helicasa Hel308 disminuyó en función de la presencia de ortovanadato de sodio o la sal de adenosina 5'-(β,γ -imido) trifosfato de litio (AMP-PNP) disminuyó la actividad de la helicasa, aumentando así el tiempo de permanencia de la helicasa. Como otro ejemplo más, tal como se describió anteriormente con referencia al Ejemplo VIII y las Figs. 17A y 17B, la concentración de una sal puede afectar al número de niveles observados. Por ejemplo, se observó que un aumento en la concentración de sal (por ejemplo, KCl) aumentaba el número de niveles observados durante la secuenciación con la helicasa Hel308 Mbu. Cualquier experto en la técnica fácilmente podría imaginar ajustes en cualquier parámetro adecuado para optimizar la manera en que se genera la señal basándose en una secuencia de polinucleótidos.

Además, debe tenerse en cuenta que las diferentes combinaciones de dichos parámetros pueden afectar la precisión de la secuenciación, así como el rendimiento de la secuenciación. Por ejemplo, al aumentar el tiempo de permanencia de la helicasa puede aumentar la precisión, por ejemplo, puede aumentar el número de niveles observados, pero potencialmente puede disminuir el rendimiento de la secuenciación. Para la secuenciación basada en la observación de etapas fraccionales, algunas etapas pueden verse afectadas más por una variable particular que por otro conjunto de etapas. Las etapas independientes de variable se pueden usar para establecer una línea de base de precisión, mientras que otras etapas se pueden ajustar para satisfacer las necesidades particulares de secuencia (por ejemplo, mayor precisión con menor rendimiento o mayor rendimiento con menor precisión). En algunas realizaciones, un dispositivo multimodal puede aprovechar esto ajustando la precisión y el rendimiento en función de las necesidades del secuenciador, por ejemplo, ajustando uno o más parámetros durante la secuenciación. Como ejemplo ilustrativo, no limitativo, y como se señaló anteriormente, se ha observado que una disminución en la concentración de ATP con Hel308 Tga puede aumentar las duraciones de los estados fraccionales. Un aumento en la duración del estado fraccional puede aumentar la precisión de la secuenciación, por ejemplo, al mejorar la relación señal/ruido (SNR) de la lectura del estado fraccional o permitir la aplicación de filtros de baja frecuencia, pero puede reducir el rendimiento. Un dispositivo multimodal puede aprovechar esto al comenzar una secuenciación con altas concentraciones de ATP para determinar de forma relativamente rápida una "estructura" aproximada de la secuencia, y luego puede reducir la concentración de ATP para "llenar los huecos" de la estructura con mayor calidad, aunque más lento, lee.

Además, téngase en cuenta que cualquier número adecuado de parámetros diferentes se puede usar de forma secuencial o en paralelo entre sí para aumentar la resolución de una o más señales producidas por la translocación por una helicasa Hel308 de un polinucleótido diana a través de un poro. Las figuras 21A-21C ilustran esquemáticamente señales que pueden generarse en función del tiempo para diferentes translocaciones de un polinucleótido a través de un poro, de acuerdo con algunas realizaciones. Las figuras 21A-21C ilustran cada una una curva discontinua, que corresponde a una señal idealizada generada en condiciones en las que un polinucleótido se

transloca a través de un poro bajo la fuerza aplicada aplicando solo una diferencia de potencial, en lugar de un motor molecular, y con una resolución de señal infinita. En tales condiciones, la señal es una función que cambia continuamente de las posiciones y la secuencia de los nucleótidos a medida que pasan a través del poro.

La Fig. 21A también ilustra una señal ejemplar (líneas gruesas) generada utilizando solo etapas de translocación completas que ocurren en los momentos indicados por las líneas de puntos verticales. La señal puede ser una señal electrónica u óptica tal como se describe en otra parte del presente documento. Además, la señal puede incluir cualquier característica adecuada de dicha señal electrónica u óptica, como el nivel de señal promedio, la duración de la señal o la desviación estándar (por ejemplo, ruido de banda ancha o ruido limitado de banda). La señal se puede ver en la fig. 21A para cambiar de un nivel relativamente bajo a un nivel relativamente alto a través de un solo paso, y luego nuevamente a un nivel relativamente bajo a través de un solo paso, que corresponde a las transiciones que ocurren entre las etapas de translocación completa a medida que el polinucleótido se transloca a través del poro. También se puede ver en la fig. 21A que la señal intersecta la señal idealizada en los puntos (a), (b) y (c) en diferentes momentos, y por lo tanto se puede considerar que "muestrea" la señal idealizada en estos puntos. Sin embargo, debido a que la tasa de muestreo efectiva es relativamente baja, la señal muestrea la señal idealizada relativamente mal. Por ejemplo, los valores en los puntos (a) y (b) son los mismos entre sí, correspondientes a niveles de señal degenerados para diferentes etapas de translocación. Debido a que la señal no muestra adecuadamente la porción de la curva idealizada que se encuentra entre los puntos (a) y (b), las etapas de translocación física correspondientes a los puntos (a) y (b) pueden ser indistinguibles entre sí, lo que resulta en la pérdida de información sobre la secuencia polinucleotídica. Además, como la señal no muestra adecuadamente la porción de la curva idealizada que se encuentra entre los puntos (b) y (c), las etapas de translocación física correspondientes a la pendiente descendente de la curva idealizada entre los puntos (b) y (c) potencialmente solo puede caracterizar parcialmente la porción del polinucleótido translocado a través del poro durante tales pasos.

Además de la señal idealizada representada por la curva discontinua como se describe anteriormente, la FIG. 21B también ilustra una señal ejemplar (líneas gruesas) generada utilizando una combinación de etapas de translocación completa separadas por el tiempo, o una combinación de etapas de translocación completa y fraccional, que ocurren en los momentos indicados por las líneas de puntos verticales. Las etapas de translocación total separadas en el tiempo pueden corresponder a señales generadas por dos motores moleculares que translocan cada uno el polinucleótido, pero en momentos que se desplazan entre sí, por ejemplo, que se desplazan entre sí aproximadamente en un 50% de duración del tiempo de un ciclo completo de translocación. Una combinación de etapas de translocación completa y fraccional puede corresponder a señales generadas por un solo motor molecular (por ejemplo, una helicasa Hel308) que transloca fraccionalmente el polinucleótido a través de etapas de translocación parcial y completa, como cuando la etapa de translocación fraccional ocurre aproximadamente a 50 % de la duración de un ciclo completo de translocación. La señal puede ser como se describe anteriormente con referencia a la FIG. 21A. La señal se puede ver en la fig. 21B para cambiar de un nivel relativamente bajo a un nivel relativamente alto a través de una secuencia de pasos, y luego nuevamente a un nivel relativamente bajo a través de otra secuencia de pasos, correspondiente a las transiciones que ocurren entre las etapas de translocación completa separadas por el tiempo o por la combinación de etapas de translocación completa y fraccional, a medida que el polinucleótido se traslada a través del poro. También se puede ver en la fig. 21B que la señal cruza la señal idealizada en un número significativamente mayor de puntos (y tiempos) que en la FIG. 21A y, por lo tanto, se puede considerar que "muestrea" la señal idealizada en estos puntos. Debido a que la tasa de muestreo efectiva es relativamente más alta que en la FIG. 21A, la señal muestrea la señal idealizada relativamente mejor que en la FIG. 21A. Por ejemplo, los valores en los puntos (a) y (b) son los mismos entre sí, correspondientes a niveles de señal degenerados para diferentes etapas de translocación. Debido a que la señal en la fig. 21A también muestrea la parte de la curva idealizada que se encuentra entre los puntos (a) y (b), las etapas de translocación física correspondientes a los puntos (a) y (b) se pueden distinguir entre sí, lo que da como resultado información adicional sobre la secuencia de polinucleótidos. Adicionalmente, debido a que la señal en la fig. 21B muestrea más completamente la porción de la curva idealizada que se encuentra entre los puntos (b) y (c), las etapas de translocación física correspondientes a la pendiente descendente de la curva idealizada entre los puntos (b) y (c) pueden caracterizar mejor la porción del polinucleótido translocado a través del poro durante tales etapas que potencialmente se pueden lograr usando la señal en la FIG. 21A.

Además de la señal idealizada representada por la curva discontinua como se describe anteriormente, la FIG. 21C también ilustra otra señal ejemplar (líneas gruesas) generada utilizando una combinación de etapas de translocación completa separadas por el tiempo, o una combinación de etapas de translocación completa y fraccional, que ocurren en los momentos indicados por las líneas de puntos verticales. Las etapas de translocación total separadas en el tiempo pueden corresponder a señales generadas por múltiples motores moleculares que translocan cada una el polinucleótido, pero en momentos que se desplazan entre sí, por ejemplo, que se desplazan entre sí aproximadamente en un 25%, 50% y 75% de la duración de un ciclo completo de translocación. Una combinación de etapas de translocación completa y fraccional puede corresponder a señales generadas por un solo motor molecular (por ejemplo, una helicasa Hel308) que transloca fraccionalmente el polinucleótido a través de etapas de translocación parcial y completa, como cuando la etapa de translocación fraccional ocurre aproximadamente en un 25%, 50% y 75% de la duración de un ciclo completo de translocación. La señal puede ser como se describe anteriormente con referencia a la FIG. 21A. La señal se puede ver en la fig. 21C que cambia de un nivel relativamente bajo a un nivel relativamente alto a través de una secuencia de un mayor número de etapas que en la FIG. 21B, y luego nuevamente a un nivel relativamente bajo a través de otra secuencia de un mayor número de etapas que en la FIG. 21B,

correspondiente a las transiciones que se producen entre las etapas de translocación completa separadas por el tiempo o por la combinación de etapas de translocación completa y fraccional, ya que el polinucleótido se transloca a través del poro. También se puede ver en la fig. 21C que la señal interseca la señal idealizada en un número significativamente mayor de puntos (y tiempos) que en la FIG. 21B y, por lo tanto, se puede considerar que "muestrea" la señal idealizada en estos puntos. Debido a que la tasa de muestreo efectiva es relativamente más alta que en la FIG. 21B, la señal muestrea la señal idealizada relativamente mejor que en la FIG. 21B y, por lo tanto, puede caracterizar mejor el polinucleótido translocado a través del poro durante tales etapas que lo que potencialmente se puede lograr utilizando la señal en la FIG. 21A o 21B.

Debe apreciarse que se puede usar cualquier selección adecuada de parámetros para aumentar el muestreo de cualquier porción seleccionada de una curva de muestra idealizada. Por ejemplo, como se mencionó anteriormente, se puede usar una combinación de etapas de translocación completa desplazadas en el tiempo (desplazadas en fase) de diferentes motores moleculares. En este sentido, aunque la fig. 21B describe el desplazamiento del tiempo de los motores moleculares en un 50% del tiempo de una etapa de translocación completa entre sí y, aunque la FIG. 21B describa el cambio de tiempo de los motores moleculares en un 25%, 50% y 75% del tiempo de una etapa de translocación completa entre sí, tales valores son puramente ilustrativos, y los motores moleculares pueden cambiarse en el tiempo en cualquier cantidad adecuada de tiempo entre sí, por ejemplo, se puede cambiar entre un 5% y un 95% del tiempo de una etapa de translocación completa entre sí, por ejemplo, se puede cambiar entre un 10% y un 90% del tiempo de una etapa de translocación completa entre sí, por ejemplo, se puede cambiar entre un 25% y un 75% del tiempo de una etapa de translocación completa, por ejemplo, se puede cambiar entre un 40% y un 60% del tiempo de una etapa de translocación completa entre sí. Como otro ejemplo, una combinación de etapas de translocación completa y fraccional puede corresponder a señales generadas por un solo motor molecular (por ejemplo, una helicasa Hel308) que transloca fraccionalmente el polinucleótido a través de etapas de translocación parcial y completa. Aunque la Fig. 21B describe las etapas de translocación fraccional que ocurren en el 50% del tiempo de una etapa de translocación completa entre sí, y aunque la FIG. 21C describa las etapas de translocación fraccional que ocurren al 25%, 50% y 75% del tiempo de una etapa de translocación completa entre sí, tales valores son puramente ilustrativos, y las etapas de translocación fraccional pueden ocurrir en cualquier momento adecuado en relación con las etapas de translocación completas, por ejemplo, del 5% al 95% del tiempo de una etapa de translocación completa, por ejemplo, del 10% al 90% del tiempo de una etapa de translocación completa, por ejemplo, del 25% al 75% del tiempo de una etapa de translocación completa, por ejemplo, del 40% al 60% del tiempo de una etapa de translocación completa.

Adicionalmente, debe apreciarse que los tiempos relativos en los que ocurren las etapas completas o fraccionales y, por lo tanto, los tiempos en los cuales la señal muestrea la señal idealizada, pueden ajustarse adecuadamente variando cualquier parámetro adecuado. Por ejemplo, como se señaló anteriormente, las variables o parámetros ejemplares que pueden afectar la generación de señales pueden incluir la temperatura, la concentración de sal (por ejemplo, Mg, Cl), la concentración del cofactor (por ejemplo, ATP), la concentración de los productos de ATP como pirofosfato, pH, el motor molecular particular utilizado y similares. En algunas realizaciones, se puede generar una primera señal basada en un primer conjunto de parámetros para muestrear la señal idealizada en un primer conjunto discreto de tiempos, y se puede generar una segunda señal basada en un segundo conjunto de parámetros (que difiere del primer conjunto de parámetros en al menos un aspecto) para muestrear la señal idealizada en un segundo conjunto discreto de veces. La primera y la segunda señales se pueden combinar para proporcionar una curva de señal que muestree la señal idealizada con mayor resolución que la primera o la segunda señal sola. Debe apreciarse que cualquier número adecuado de señales puede combinarse de manera análoga para proporcionar una curva de señal que muestree la señal idealizada con mayor resolución que cualquier otra señal individual.

Ejemplo X

Enfoques adicionales para la identificación de secuencias

Algunos enfoques adicionales para la identificación de secuencias se describen con referencia al Ejemplo X.

Ciertos tipos de información pueden usarse solos, o en combinación entre sí, para obtener información específica de la secuencia: (A) información de respuesta de etapa completa solo, (B) información de respuesta de etapa fraccional solo, (C) información de respuesta etapa completa y de etapas fraccionales juntas sin identificadores, y (D) información de respuesta de etapas fraccionales y etapas completas junto con identificadores.

Por "información de respuesta" se entiende los datos obtenidos de la respuesta del sistema frente a una secuencia de polinucleótidos dada (k-mer) que es exclusiva de ese k-mer o un subconjunto de k-mers (incluido el k-mer de interés). Los ejemplos de información de respuesta incluyen una corriente de nivel medio, corriente de nivel medio, ruido de corriente de nivel de banda ancha, ruido de corriente de nivel limitado de banda, duración de nivel y similares.

Por "identificadores" se entiende los datos obtenidos mientras que el polinucleótido (k-mer) interactúa con el entorno de los nanoporos que identifica dónde a lo largo de la "respuesta idealizada" se encuentra un nivel particular con respecto a otros niveles. Por ejemplo, los sistemas que utilizan la helicasa Hel308 Tga como un motor molecular en presencia de niveles relativamente altos o relativamente bajos de concentración de ATP pueden mostrar duraciones relativamente cortas o relativamente largas, respectivamente, para cualquier otro nivel, donde todos los demás niveles son aproximadamente el 50% de la respuesta idealizada de los niveles vecinos. En este ejemplo, la duración del nivel

se puede usar como un identificador, porque se puede usar para identificar la ubicación de la secuencia a lo largo de la respuesta idealizada (con respecto a los niveles vecinos).

Por "respuesta idealizada" se entiende la respuesta del sistema para un polinucleótido particular que se transloca a través del nanoporo con una resolución suficientemente alta de modo que se puedan resolver los movimientos del polinucleótido lo suficientemente pequeños. Por ejemplo, una respuesta idealizada es una traza de corriente continua de resolución infinitamente alta de ADN que se transloca a través del nanoporo.

Refiriéndose nuevamente a los elementos (A) - (D) mencionados más arriba en este ejemplo, cada uno de los elementos (A) - (D) se puede utilizar de forma independiente, o junto con uno o más de los otros elementos (A) - (D) para identificar una secuencia polinucleotídica. Por ejemplo, uno o más de los ítems (A) - (D) se pueden calcular independientemente a partir de cualquier otro ítem (A) - (D), por ejemplo, debido a las restricciones de recursos computacionales, las restricciones de tiempo, el conocimiento a priori de un enfoque óptimo, y similares. Sobre la base de que se calcule más de uno de los ítems (A) - (D), se pueden utilizar los resultados de solo uno de los ítems (A) - (D). La determinación de cuál de estos cálculos utilizar puede basarse en la confianza en los resultados. Por ejemplo, la confianza en los resultados se puede basar en uno o más de los siguientes: (a) la información de respuesta en sí misma (por ejemplo, los niveles altos de ATP pueden acortar los tamaños de etapas fraccionales en Hel308 Tga, lo que puede reducir la confianza en el elemento (B) relativo al artículo (A)); (b) el propio algoritmo de secuenciación (por ejemplo, un algoritmo Viterbi puede producir una puntuación de probabilidad para la secuencia óptima que propone, que se puede usar para determinar un nivel de confianza en esa secuencia propuesta); (c) la secuencia producida por el algoritmo de secuenciación (por ejemplo, la confianza puede asignarse en función de una comparación entre la secuencia propuesta por el algoritmo y una tabla de búsqueda de secuencias y/o cualquier conocimiento a priori del polinucleótido que se está secuenciando); o (d) cualquier combinación adecuada de elementos (a) - (c).

Téngase en cuenta que, en algunas circunstancias, puede ser beneficioso determinar la secuencia real utilizando las secuencias propuestas de más de uno de los ítems (A) - (D). Por ejemplo, se podría determinar una secuencia de consenso basada en algunas o todas las secuencias propuestas. La secuencia de consenso se puede determinar en base a todas o algunas de las secuencias propuestas. La secuencia de consenso se puede aplicar globalmente a toda la secuencia de polinucleótidos o localmente a una parte de la secuencia. La secuencia de consenso se puede determinar en función de los valores de confianza de algunos o todos los ítems (A) - (D). Los valores de confianza pueden ser los descritos más arriba en este ejemplo. Los valores de confianza se pueden aplicar localmente, a una parte de una secuencia, o globalmente, a toda la secuencia. Una secuencia de consenso final se puede determinar mediante múltiples rondas de los enfoques mencionados anteriormente, donde el consenso resultante de cada ronda se puede usar como una secuencia propuesta, y los métodos de determinación de confianza para cada ronda pueden diferir entre rondas.

Como ejemplo, se puede usar un algoritmo de Viterbi para determinar dos secuencias propuestas diferentes mediante la secuenciación de solo etapas completas y solo etapas fraccionales de translocación de ADN a través de un nanoporo (elementos (A) y (B) descritos anteriormente en este ejemplo). Las puntuaciones de probabilidad de este algoritmo para cada fragmento de ADN se utilizan para determinar la confianza para cada región de la secuencia propuesta, y una agregación de confianzas para cada región puede dar como resultado una secuencia de consenso propuesta de primera ronda. Esta secuencia de consenso puede compararse con las dos secuencias propuestas inicialmente con respecto a una tabla de búsqueda de secuencias conocidas. La similitud entre la tabla de consulta y cada una de estas tres secuencias propuestas puede dar como resultado valores de confianza para cada región de cada una de las tres secuencias propuestas. Esta segunda ronda de comparaciones basadas en la confianza entre las tres secuencias propuestas puede dar como resultado una secuencia de consenso final propuesta.

En algunas realizaciones, las figs. 22A-22D ilustran etapas en métodos ilustrativos para usar la información proporcionada por la translocación fraccional de un polinucleótido a través de un poro, de acuerdo con algunas realizaciones. La Fig. 22A ilustra una descripción general de alto nivel de un método para usar información proporcionada por la translocación fraccional de un polinucleótido a través de un poro, de acuerdo con algunas realizaciones. El método ilustrado en la fig. 22A incluye obtener una señal (etapa 2210), tal como una o más señales producidas por una o más etapas de translocación fraccional por una helicasa Hel308 de un polinucleótido diana a través de un poro tal como se describe con mayor detalle en otra parte en este documento. El método ilustrado en la fig. 22A también incluye la detección e identificación del nivel (etapa 2220), por ejemplo, la detección e identificación de diferentes niveles de señal en la señal, por ejemplo, la detección e identificación de niveles que corresponden a etapas de translocación fraccional del polinucleótido a través del poro, y también la detección e identificación de los niveles que corresponden a las etapas completas de translocación del polinucleótido a través del poro. El método ilustrado en la fig. 22A también incluye la determinación de secuencia (etapa 2230), por ejemplo, caracterizando la secuencia del polinucleótido en base a los diferentes niveles de señal detectados e identificados en la señal. El método ilustrado en la fig. 22A también incluye la resolución de una secuencia (etapa 2240), por ejemplo, la resolución de una secuencia probable de nucleótidos del nucleótido real en base a los resultados de la resolución de secuencia.

Las figuras 22B-22D ilustran subpasos opcionales de una o más de las etapas ilustradas en la FIG. 22A. Por ejemplo, la fig. 22B ilustra detalles adicionales de una implementación potencial de las etapas 2210 y 2220 ilustradas en la FIG. 22A. El método ilustrado en la fig. 22B nuevamente incluye la obtención de una señal (etapa 2210), como una o más

señales producidas por una o más etapas de translocación fraccional por una helicasa Hel308 de un polinucleótido diana a través de un poro tal como se describe con mayor detalle en otra parte en este documento. El método ilustrado en la fig. 22B también puede incluir opcionalmente obtener parámetros de entrada (2211). Dichos parámetros de entrada pueden incluir, entre otros, parámetros que definan qué características de la señal ilustrativas deben detectarse y determinarse para que correspondan a una señal. Por ejemplo, los parámetros de entrada pueden definir un cambio de magnitud del umbral en el valor de la señal, por encima del cual un cambio de magnitud en la señal se puede detectar como correspondiente a un nivel. O, por ejemplo, los parámetros de entrada pueden definir que solo se deben detectar los niveles de señal correspondientes a las etapas de translocación completa, o solo los niveles de señal correspondientes a las etapas de translocación fraccional, o los niveles de señal correspondientes a las etapas de translocación completa y fraccional. Los parámetros de entrada también pueden contener información asociada con los modos de error (por ejemplo, omisión de nucleótidos o alternancia de nucleótidos), posiblemente incluyendo la propensión y/o el grado de ciertos modos de error, que se pueden tener en cuenta al determinar los niveles. Los parámetros de entrada también pueden incluir información asociada con el entorno particular entre el que interactúan el nanoporo, el motor molecular y el polinucleótido (por ejemplo, la temperatura, la salinidad, el pH, la concentración del cofactor, etc.), que se pueden usar para determinar los niveles para una señal dada. El método ilustrado en la fig. 22B también incluye la detección del nivel, por ejemplo, la detección de diferentes niveles de señal en la señal que corresponden a etapas de translocación fraccional del polinucleótido a través del poro (paso 2221). Por ejemplo, basándose en la señal obtenida en la etapa 2210 y los parámetros de entrada obtenidos en la etapa 2211, dicha detección de nivel puede detectar regiones de la señal que son estadísticamente significativamente diferentes de otras regiones de la señal como para corresponder al nivel. Los métodos ejemplares de detección de nivel (que también se pueden denominar detección de bordes o detección de pasos) son conocidos en la técnica e incluyen la prueba t de Student y la maximización de chi-cuadrado. Para algunos ejemplos de algoritmos de detección de etapas que pueden adaptarse adecuadamente para su uso en la detección de niveles en la etapa 2221, véase Carter et al., "A Comparison of Step-Detection Methods: How Well Can You Do", *Biophysical Journal* 94: 306 -308 (enero 2008).

El método ilustrado en la fig. 22B también incluye información de nivel de resolución (paso 2222) basada en la detección de nivel de la etapa 2221. La información de nivel puede incluir el promedio, la mediana, el modo, la distribución, la duración, el máximo y/o la corriente mínima detectada para un nivel determinado, o cualquier combinación de estos valores, o estos valores pertenecientes a un subconjunto de valores de corriente para un nivel determinado (por ejemplo, se puede utilizar la corriente promedio después de eliminar primero la información de corriente asociada con los modos de error). La información de nivel también puede incluir la desviación estándar de la corriente, o un subconjunto de corriente de banda limitada de frecuencia (por ejemplo, la corriente obtenida después de aplicar un filtro a paso bajo, paso alto, paso de banda o parada de banda, o cualquiera combinación de estos filtros). La información de nivel también puede incluir información asociada con la duración de los niveles, así como información de modo de error asociada con los niveles. El método ilustrado en la fig. 22B también incluye la identificación de nivel (paso 2223), por ejemplo, determinar cuál de los niveles detectados en la etapa 2221 para qué información de nivel se envía en la etapa 2222 corresponde a las etapas de translocación total o fraccional del polinucleótido diana. Por ejemplo, la etapa 2223 puede incluir analizar las duraciones de los diferentes niveles detectados en la etapa 2221 para los cuales se entrega información de nivel en la etapa 2222, y en base a dichas duraciones, identificar ciertos niveles como correspondientes a las etapas completas de translocación e identificar otros niveles determinados como correspondientes a las etapas de translocación fraccional. Como ejemplo, se puede suponer que los niveles de señal que tienen una duración más corta que un primer umbral corresponden a ruido y, por lo tanto, se descartan, mientras que los niveles de señal que tienen una duración mayor que un primer umbral y más cortos que un segundo umbral pueden suponer que corresponden a una etapa de translocación fraccional y, por lo tanto, ser identificado como tal, mientras que los niveles de señal que tienen una duración más larga que el segundo umbral y más corta que un tercer umbral pueden suponer que corresponden a una etapa de translocación completa y, por lo tanto, se identifican como tales, mientras que los niveles de señal que tienen una duración mayor que la se puede suponer que el tercer umbral corresponde a un error, o a una ausencia de polinucleótido y, por lo tanto, se descarta.

El método ilustrado en la fig. 22B también incluye la resolución de una o más de las siguientes resoluciones: niveles completos, niveles fraccionales, todos los niveles e identificadores de nivel. Por ejemplo, como se indicó anteriormente, los parámetros de entrada obtenidos en la etapa 2211 pueden definir que solo los niveles de señal correspondientes a las etapas de translocación completa, o solo los niveles de señal correspondientes a las etapas de translocación fraccional, o los niveles de señal correspondientes a las etapas de translocación completa y fraccional (por ejemplo, "todos los niveles"), deben ser detectados. Téngase en cuenta que, en algunas realizaciones, la selección de "todos los niveles" a través de los parámetros de entrada puede corresponder a omitir la etapa de identificación de nivel, de manera que la etapa de detección de nivel 2221 genera directamente todos los niveles. Alternativamente, en función de los resultados de la detección de nivel 2223 y los parámetros de entrada 2211, los niveles identificados de las señales deseadas pueden emitirse, por ejemplo, para un procesamiento adicional como el que se describe a continuación con referencia a las Figs. 22C y 22D. Los identificadores de nivel pueden incluir cualquier información adecuada que facilite un análisis adicional de los niveles, por ejemplo, índices que indiquen las duraciones de las etapas completas o fraccionales que se usaron durante la etapa 2223 para indicar el tipo de transición al que corresponde un nivel identificado.

Refiriéndose nuevamente a la FIG. 22A, uno o más de los niveles completos, los niveles fraccionales, todos los niveles y los identificadores de nivel, que se pueden generar utilizando el método ilustrado en la FIG. 22B o utilizando otro

método adecuado, se puede usar como entrada para realizar la determinación de secuencia (paso 2230 en la FIG. 22A). Por ejemplo, la fig. 22C ilustra un primer método ejemplar para realizar la determinación de secuencia basada en uno o más de tales niveles completos, niveles fraccionales, todos los niveles e identificadores de nivel, por ejemplo, tomando como entrada uno o más de niveles completos, niveles fraccionales, todos los niveles e identificadores de nivel. El método ilustrado en la fig. 22C incluye una etapa de resolución de secuencia basada en la entrada de uno o más de niveles completos, niveles fraccionales, todos los niveles e identificadores de nivel (paso 2231). La resolución de secuencia puede incluir cualquier método adecuado basado en qué bases de nucleótidos del polinucleótido diana pueden resolverse en función de los niveles de la señal de entrada. Los métodos ejemplares para la resolución de secuencia incluyen, pero no se limitan a, algoritmos de Viterbi tales como los descritos en el Ejemplo VII con referencia a la FIG. 19A, algoritmos de Viterbi modificados, tal como se describen en el Ejemplo VII con referencia a la FIG. 19B, o coincidencia de patrón análoga a la descrita en el Ejemplo XI. Se pueden usar otros métodos para la secuenciación de resoluciones adecuadamente. La realización de la resolución de la secuencia (paso 2231) puede incluir una pluralidad de secuencias resueltas, por ejemplo, Secuencia A, Secuencia B, ... Secuencia N, así como información de confianza para cada secuencia resuelta. Las diferentes secuencias resueltas pueden basarse en una entrada diferente a la etapa 2231 que otra. Por ejemplo, una primera secuencia resuelta (por ejemplo, la secuencia A) se puede basar en una entrada a la etapa 2231 en la que solo se identifican niveles de translocación completos en base a una señal dada obtenida en la etapa 2210, una segunda secuencia resuelta (por ejemplo, la secuencia B) puede basarse en una entrada a la etapa 2231 en la que solo se identifican los niveles de translocación fraccional, y una tercera secuencia resuelta (por ejemplo, la secuencia N) puede basarse en una entrada a la etapa 2231 en la que se identifican todos los niveles de translocación (por ejemplo, tanto niveles de translocación completos como fraccionales). Alternativamente, o adicionalmente, otras secuencias resueltas pueden basarse en otros niveles que se identificaron en base a parámetros de entrada alternativos obtenidos en la etapa 2211, tales como diferentes valores de parámetros que definen qué características características de la señal se deben detectar y determinar para que correspondan a una señal, tal como diferentes cambios de magnitud de umbral en el valor de la señal, por encima del cual se puede detectar un cambio de magnitud en la señal como correspondiente a un nivel. Cada secuencia resuelta diferente puede tener información de confianza asociada, por ejemplo, un valor que represente la probabilidad de que la secuencia resuelta se corresponda con la secuencia real del nucleótido diana.

En la realización ilustrada en la fig. 22C, una etapa de selección de secuencia (paso 2232) puede seleccionar una o más de las secuencias resueltas y proporcionar la secuencia seleccionada como resolución (paso 2240). Como ejemplo, la etapa de selección de secuencia (paso 2232) puede incluir la comparación de la información de confianza para las diversas secuencias resueltas, y puede seleccionar y emitir en la etapa 2240 la secuencia resuelta que tenga la mayor confianza, por ejemplo, la mayor probabilidad de corresponder con la secuencia real. Como otro ejemplo, la información de confianza para una secuencia resuelta dada puede incluir una pluralidad de valores de confianza que representen respectivamente la probabilidad de que las porciones correspondientes de la secuencia resuelta se correspondan con la secuencia real del polinucleótido diana para esa porción. Para diferentes porciones de las secuencias resueltas (p. ej., porciones que tienen 10 pares de bases de longitud, o 50 pares de bases de longitud, o 100 pares de bases de longitud, o 10-100 pares de bases de longitud, o 10-50 pares de bases de longitud, o 50-100 pares de bases de longitud), la etapa de selección de secuencia (paso 2232) puede incluir comparar el valor de confianza para diferentes secuencias resueltas en esa parte, y seleccionar la parte de la secuencia resuelta que tenga el valor más alto para esa parte. Esa porción seleccionada puede concatenarse con las porciones seleccionadas de otras secuencias resueltas que tienen el valor más alto para dichas porciones, o pueden alinearse con ésta.

La Fig. 22D ilustra un método alternativo que se puede usar para la determinación de secuencias (2230). El método ilustrado en la fig. 22D puede incluir obtener como entrada una pluralidad de secuencias resueltas, por ejemplo, Secuencia A, Secuencia B, ... Secuencia N, así como la información de confianza para cada secuencia resuelta, que puede ser análoga a las descritas anteriormente con referencia a la FIG. 22C. A este respecto, aunque no se ilustra específicamente, el método ilustrado en la FIG. 22D puede incluir la etapa 2231 de la secuencia de resolución que recibe entradas análogas a las descritas anteriormente con referencia a la FIG. 22C, proporciona resoluciones análogas a las descritas anteriormente con referencia a la FIG. 22C, y funciona de forma análoga a la etapa 2231. Alternativamente, el método ilustrado en la fig. 22D puede obtener una pluralidad de secuencias resueltas de cualquier otra fuente adecuada.

El método ilustrado en la fig. 22D también puede incluir obtener secuencias modelo (paso 2234). Por ejemplo, tales secuencias pueden incluir secuencias conocidas a priori para una o más especies diferentes, tales como uno o más patógenos diferentes. De manera ilustrativa, las secuencias modelo pueden almacenarse en una tabla de búsqueda, base de datos u otra estructura de datos adecuada almacenada en un medio legible por computadora no transitorio. El método ilustrado en la fig. 22D también puede incluir una etapa de selección de secuencia (paso 2233). En la realización ilustrada en la fig. 22D, la etapa de selección de secuencia puede seleccionar una o más de las secuencias resueltas recibidas como entrada, basándose en una o más de las secuencias modelo obtenidas en la etapa 2234, y proporcionar como resolución una secuencia propuesta y una nueva información de confianza. Como ejemplo, la etapa de selección de secuencia (etapa 2233 en la Figura 22D) puede incluir comparar una o más de las diversas secuencias resueltas con una o más de las secuencias modelo obtenidas en la etapa 2234, y puede seleccionar y generar una secuencia propuesta, que puede corresponder a la secuencia resuelta que tiene la información de confianza nueva más alta, por ejemplo, la probabilidad más alta de corresponder con la secuencia modelo. La información de confianza de entrada se puede sopesar con la probabilidad de que una secuencia (o regiones dentro de esa secuencia) coincida

con una secuencia modelo (o regiones dentro de la secuencia modelo) para determinar la secuencia más probable, que se puede mostrar como la secuencia propuesta. Por ejemplo, para la secuencia de entrada A la mejor alineación con la secuencia del modelo Z y la secuencia de entrada B la mejor alineación con la secuencia modelo Y, una secuencia propuesta puede ser la secuencia modelo Z con la mejor alineación entre A y Z que entre B e Y. Sin embargo, en los casos donde B e Y tienen una mejor alineación, los casos en que A tiene un valor de confianza más alto que B pueden permitir que Z sea la secuencia propuesta. Además, en otro escenario, se pueden comparar regiones de secuencias, permitiendo que la secuencia propuesta de resolución incluya información de secuencia de A, B, Z e Y. Alternativamente, una secuencia resuelta dada puede incluir nueva información de confianza, por ejemplo, una pluralidad de los nuevos valores de confianza representan respectivamente la probabilidad de que las partes correspondientes de la secuencia resuelta correspondan a partes de una o más secuencias del modelo para esa parte. Para diferentes porciones de las secuencias resueltas (p. ej., porciones que tienen 10 pares de bases de longitud, o 50 pares de bases de longitud, o 100 pares de bases de longitud, o 10-100 pares de bases de longitud, o 10-50 pares de bases de longitud, o 50-100 pares de bases de longitud), la etapa de selección de secuencia (paso 2233) puede incluir comparar el nuevo valor de confianza para diferentes secuencias resueltas en esa porción para la secuencia o secuencias modelo, y seleccionar la porción de la secuencia resuelta que tenga el mayor valor de confianza nuevo para esa porción. Esa porción seleccionada puede concatenarse con las porciones seleccionadas de otras secuencias resueltas que tienen el valor de confianza nuevo más alto para dichas porciones, o puede alinearse con ésta.

El método ilustrado en la fig. 22D puede incluir además determinar, basándose en la nueva información de confianza que se genera en la etapa 2233, si la nueva información de confianza para la secuencia propuesta, también obtenida en la etapa 2233, cumple con los requisitos (paso 2235). Como ejemplo, la etapa 2235 puede comparar la nueva información de confianza, que puede ser un nuevo valor de confianza, con un valor de confianza umbral exacto o por encima del cual se puede determinar la secuencia propuesta para que coincida suficientemente con el modelo, y por debajo del cual la secuencia propuesta puede ser determinada para no coincidir suficientemente con el modelo. La nueva información de confianza puede incluir el resultado de la información de confirmación de entrada, la relación entre la secuencia propuesta y las secuencias de entrada, la relación entre la secuencia propuesta y las secuencias modelo, y/o la relación entre las secuencias de entrada y las secuencias modelo. Por ejemplo, en el caso de que la secuencia propuesta sea simplemente una de las secuencias de entrada, la nueva información de confianza puede ser una media ponderada entre el valor de confianza de entrada de la secuencia de entrada y su puntuación de alineación para la secuencia de modelo de mejor alineación. En otros casos, como cuando la secuencia propuesta es una combinación de regiones de secuencias de entrada, la nueva información de confianza puede incluir una media ponderada de las medias ponderadas de los valores de confianza de entrada y las puntuaciones de alineación (para las secuencias modelo) entre las regiones de la secuencia propuesta. Basado en la determinación en la etapa 2235 de que la nueva información de confianza cumple con los requisitos ("sí"), la etapa 2235 proporciona la secuencia propuesta como resolución (etapa 2240). Basado en la determinación en la etapa 2235 de que la nueva información de confianza no cumple con los requisitos, la etapa 2235 regresa a la etapa 2233 en la cual continúa la selección de secuencia, por ejemplo, realizando comparaciones adicionales de las secuencias resueltas para modelar secuencias. El algoritmo de selección de secuencia o el conjunto de secuencias modelo pueden depender de los parámetros, que pueden incluir uno o más de la secuencia propuesta, nueva información de confianza, el número de veces que se ha ejecutado el algoritmo de selección de secuencia y las secuencias modelo que ya se han utilizado. Por ejemplo, una etapa inicial a través del algoritmo de selección de secuencia puede utilizar relativamente pocas secuencias modelo (por ejemplo, por el bien del rendimiento). Sin embargo, si la alineación entre las secuencias de entrada y las secuencias modelo es relativamente pobre, entonces la nueva información de confianza puede no cumplir los requisitos, y por lo tanto, la comparación con un conjunto nuevo o más refinado de secuencias modelo puede realizarse al regresar a la etapa 2233.

Ejemplo XI

Reconocimiento de patrones, opcionalmente para la identificación de SNP

En algunas realizaciones, los métodos y composiciones descritos en el presente documento pueden usarse en combinación con métodos para la detección, genotipado y amplificación de ácidos nucleicos múltiplex. Los métodos para la detección, genotipado y amplificación de ácidos nucleicos múltiplex son conocidos en la técnica y pueden ser seleccionados y aplicados fácilmente por cualquier persona con experiencia ordinaria. Por ejemplo, en una realización, los métodos y composiciones descritos en el presente documento se pueden usar en combinación con los métodos de detección, genotipado y amplificación de ácidos nucleicos múltiplex descritos en las patentes de EE.UU. Nos. 6.890.741, 6.913.884, 7.955.794, 7.582.420 y 8.288.103, y la Publicación de EE.UU. 2013-0244882.

En algunas realizaciones, los métodos para la detección, genotipado y amplificación de ácidos nucleicos múltiplex que pueden combinarse con los métodos y composiciones descritos en el presente documento incluyen métodos realizados en combinación con un soporte sólido, como una matriz (aleatoria y ordenada) o perlas. Por ejemplo, en algunos aspectos, los polinucleótidos diana a ensayar, como el ADN genómico, pueden inmovilizarse en un soporte sólido. Dichos polinucleótidos diana inmovilizados pueden someterse a los métodos de genotipado y detención de ácidos nucleicos múltiplex que son conocidos en la técnica. El polinucleótido diana resultante se puede caracterizar usando los métodos descritos en el presente documento.

En algunas realizaciones, los métodos para caracterizar un polinucleótido diana pueden incluir además las etapas

necesarias para generar el polinucleótido diana a ensayar. Por consiguiente, en algunas realizaciones, el método puede incluir las etapas de: (a) proporcionar una pluralidad de secuencias de ácido nucleico diana que comprenden cada una de 3' a 5' de un primer, segundo y tercer dominio diana, comprendiendo el primer dominio diana una posición de detección siendo el segundo dominio diana al menos un nucleótido; (b) poner en contacto las secuencias de ácido nucleico diana con conjuntos de sondas para cada secuencia diana para formar un conjunto de primeros complejos de hibridación, comprendiendo cada conjunto de sondas: una primera sonda que comprende de 5' a 3', una secuencia de cebado universal, y una secuencia sustancialmente complementaria al primer dominio diana de una secuencia diana, y una posición de interrogación adecuada para la base con la posición de detección (por ejemplo, dentro de las cuatro bases terminales 3'), y una segunda sonda que comprende de 5' a 3', una secuencia sustancialmente complementaria al tercer dominio diana de una secuencia diana y una secuencia de cebado universal, en donde opcionalmente al menos una sonda contiene una secuencia de identificación de locus (por ejemplo, marcador o señuelo); (c) poner en contacto los complejos de hibridación con una enzima de extensión y dNTP, en condiciones en las que si la base en las posiciones de interrogación es perfectamente complementaria con las bases en las posiciones de detección, la extensión de las primeras sondas se realiza a través de los segundos dominios de destino para formar segundos complejos de hibridación y (d) unir las primeras sondas extendidas a las segundas sondas para formar plantillas de amplificación. En algunos aspectos de este método, la primera o la segunda sonda de los conjuntos de sondas puede incluir una secuencia de identificación de alelo (por ejemplo, un marcador o un señuelo).

En algunas realizaciones, los métodos para caracterizar un polinucleótido diana pueden incluir además las etapas de: (a) proporcionar una pluralidad de secuencias de ácido nucleico diana que comprenden cada una de 3' a 5' un primer, segundo y tercer dominio diana, comprendiendo el primer dominio diana una posición de detección, siendo el segundo dominio diana al menos un nucleótido; (b) poner en contacto las secuencias de ácido nucleico diana con sondas que comprenden cada una de 5' a 3', una secuencia de cebado universal y una secuencia sustancialmente complementaria al primer dominio de destino de una secuencia diana, y una posición de interrogación adecuada para el cotejo de bases con la posición de detección (por ejemplo, dentro de las cuatro bases terminales 3'), en donde opcionalmente las sondas contienen una secuencia de identificación de locus (por ejemplo, marcador o señuelo); (c) poner en contacto los complejos de hibridación con una enzima de extensión y los dNTP, en condiciones en las que si la base en las posiciones de interrogación son perfectamente complementarias con las bases en las posiciones de detección, la extensión de las sondas se realiza a través del segundo y tercer dominio de destino para formar sondas extendidas que pueden actuar como plantillas de amplificación.

El método para generar el polinucleótido diana para ensayar en los métodos descritos en el presente documento puede incluir además amplificar las plantillas de amplificación para producir amplicones. En algunos aspectos, los cebadores que comprenden la secuencia de cebado universal para la primera o segunda sonda también incluyen una secuencia de identificación de alelo o una secuencia de identificación de locus (por ejemplo, marcador o señuelo), dependiendo de qué secuencia de identificación ya se haya incorporado en la plantilla de amplificación. Estos amplicones, que pueden incluir tanto una secuencia de identificación de locus como una secuencia de identificación de alelos, pueden caracterizarse usando los métodos descritos en este documento. La caracterización de la secuencia diana puede indicar el genotipo de la muestra en función de la presencia del locus y las secuencias de identificación de alelos.

En algunas realizaciones, los cebadores utilizados para producir amplicones incluyen uno o más residuos modificados que no permiten que la enzima de extensión utilizada durante la amplificación atraviese los residuos. Por ejemplo, en algunos aspectos, un cebador incluye un sitio abásico (sitioapurínico/apirimidinico), un fosoramidita espaciador C3 (espaciador C3 Int), un espaciador de trietilenglicol (espaciador Int 9) o un espaciador hexa-etilenglicol de 18 átomos (espaciador Int 18) para evitar que la enzima de extensión continúe la extensión del cebador. Se entiende que cualquier persona con experiencia ordinaria en la técnica puede seleccionar otros residuos modificados que pueden realizar esta misma función. El uno o más residuos modificados pueden ubicarse dentro de la secuencia de identificación del alelo o en cualquiera de los lados de la secuencia de identificación del alelo siempre que se genere un saliente de 5' de longitud suficiente para caracterizar el polinucleótido diana utilizando los métodos descritos en este documento. Por ejemplo, el saliente de 5' tiene una longitud suficiente para permitir la inmovilización del amplicón.

En algunas realizaciones, los amplicones generados por los métodos anteriores se ponen en contacto adicionalmente con una endonucleasa de corte para generar un saliente de 3' en la segunda secuencia de la sonda o cerca de ésta. Dichas enzimas de corte pueden ser específicas de la secuencia, de modo que solo una hebra de un producto de doble cadena se escinde. Una variedad de endonucleasas de corte son conocidas en la técnica y se reconoce que cualquier persona con experiencia ordinaria puede seleccionar fácilmente la endonucleasa apropiada basándose en la sonda y la secuencia de cebado. Con el fin de generar el saliente 3' después de la escisión por la endonucleasa de corte, se pueden usar varios métodos conocidos en la técnica que incluyen, por ejemplo, desnaturizar parcialmente los amplicones de modo que la porción más pequeña de la cadena cortada se libere del amplicón, mientras que el resto del amplicón permanezca hibridado junto. Para facilitar la eliminación de la porción más pequeña del amplicón, se puede agregar un complemento inverso de la porción más pequeña para hibridarse con la cadena no deseada.

En algunas realizaciones, se puede generar un saliente de 3' incluyendo uno o más residuos de uracilo en la segunda secuencia de la sonda descrita en los métodos anteriores y poniendo en contacto el amplicón con una enzima específica de uracilo que genere específicamente un único espacio de nucleótidos en la ubicación del uracilo. Un ejemplo no limitativo de dicha enzima específica de uracilo es la enzima reactiva de escisión específica de uracilo

(USER™) (New England Biolabs). Por consiguiente, los fragmentos intercalados más pequeños generados por la enzima pueden desnaturalizarse fácilmente desde el amplicón utilizando métodos conocidos.

En ciertos aspectos, el saliente de 3' que se genera es de una longitud suficiente para facilitar la unión de una helicasa descrita en el presente documento. En consecuencia, en algunos aspectos, el saliente 3' incluye al menos 4 nucleótidos de longitud. En otros aspectos, el saliente 3' incluye entre 4-20 nucleótidos de longitud, o en ciertos aspectos entre 8-16, o en otros aspectos entre 10 y 16 nucleótidos de longitud.

La frase "secuencia de identificación de locus" se refiere a una secuencia de residuos de ácido nucleico (por ejemplo, un marcador o señuelo) que se ha asignado o se sabe que está conectada a una ubicación particular en un polinucleótido diana. La ubicación de un polinucleótido diana puede ser, por ejemplo, un gen, una porción de un gen (por ejemplo, un exón o intrón) o una región no codificante (por ejemplo, un promotor o un potenciador) en un genoma que está cerca del alelo que está siendo ensayado. La secuencia de identificación de locus puede ser una secuencia natural que sea específica para la ubicación de la secuencia de interés diana y/o una secuencia sintética que no sea nativa de la secuencia de interés diana. La secuencia de identificación del locus puede asignarse mediante un patrón de señal esperado del marcador o señuelo.

La frase "secuencia de identificación de alelo" se refiere a una secuencia de residuos de ácido nucleico (por ejemplo, un marcador o señuelo) que se ha asignado a un residuo de ácido nucleico específico que se encuentra en una posición de detección de un polinucleótido diana. La secuencia de identificación del alelo puede indicar la presencia de un residuo de ácido nucleico (por ejemplo, A, T, C o G) en una posición de detección. La secuencia de identificación del alelo también puede asignarse mediante un patrón de señal esperado del marcador o el señuelo.

En otra realización, los métodos para caracterizar un polinucleótido diana pueden incluir además las etapas descritas en la FIG. 18. Dicho método puede incluir las etapas de: (a) proporcionar una muestra que tenga diferentes secuencias de ácido nucleico diana de interés, en donde las diferentes secuencias de ácido nucleico diana están opcionalmente inmovilizadas sobre un soporte sólido; (b) poner en contacto la muestra con un conjunto de sondas para cada una de las diferentes secuencias de ácido nucleico diana de interés para formar complejos de hibridación, comprendiendo cada conjunto: una primera sonda que comprende de 5' a 3': una primera secuencia de cebado universal y una secuencia que sea sustancialmente complementaria al primer dominio de destino y que tenga una posición de interrogación adecuada para el cotejo con la posición de detección; y una segunda sonda que comprenda de 5' a 3': una secuencia sustancialmente complementaria al tercer dominio de destino, y una segunda secuencia de cebado universal, en donde al menos una sonda contiene una secuencia de identificación de locus (por ejemplo, marcador o señuelo) que no es original de la secuencia diana de interés; (c) poner en contacto los complejos de hibridación con una enzima de extensión y dNTP, en donde para cada complejo de hibridación, si la base en la posición de interrogación es perfectamente complementaria a la base en la posición de detección, entonces la primera sonda se extiende a lo largo del segundo dominio de destino; (d) unir las primeras sondas extendidas a las segundas sondas para formar plantillas de amplificación; (e) amplificar las plantillas de amplificación con el primer y segundo cebadores universales para producir amplicones, en donde al menos un cebador incluye una secuencia de identificación de alelo (por ejemplo, marcador o señuelo), en donde la secuencia de identificación de alelo comprende un sitio abásico; (f) poner en contacto los amplicones con una endonucleasa de corte para generar un saliente de 3' en la segunda secuencia del cebador; y (g) detectar la presencia tanto de la secuencia de identificación de locus como de la secuencia de identificación de alelos de diferentes amplicones utilizando los métodos para caracterizaciones de polinucleótidos diana descritos en el presente documento, indicando así la presencia de las diferentes secuencias de interés diana en la muestra.

Como se usa en el presente documento, la frase "múltiplex", o sus equivalentes gramaticales, se refiere a la detección, análisis o amplificación de más de una secuencia diana de interés. En una realización, el múltiplex se refiere a al menos 100 ó 200 secuencias diana diferentes, mientras que se prefieren al menos 500 secuencias diana diferentes. Más preferido es al menos 1000, siendo más de 5000 o 10.000 particularmente preferidos y más de 50.000 ó 100.000 los más preferidos. La detección se puede realizar en una variedad de plataformas como se describe en este documento.

En algunos aspectos, la descripción en este documento proporciona métodos para la detección de secuencias diana de ácido nucleico en una muestra. Como apreciarán los expertos en la técnica, la solución de muestra puede comprender cualquier cantidad de sustancias, incluidos, entre otros, fluidos corporales (incluidos, entre otros, sangre, orina, suero, secreciones linfáticas, salivares, anales y vaginales, transpiraciones y semen, de prácticamente cualquier organismo, prefiriéndose las muestras de mamíferos y, particularmente, las muestras humanas; muestras ambientales (incluidas, entre otras, muestras de aire, agricultura, agua y suelo); muestras de agentes de guerra biológica; muestras de investigación; muestras purificadas, como ADN genómico purificado, ARN, proteínas, etc.; muestras crudas (bacterias, virus, ADN genómico, etc.). Como apreciarán los expertos en la técnica, prácticamente cualquier manipulación experimental puede haberse realizado en la muestra.

Si es necesario, el polinucleótido diana se prepara utilizando técnicas conocidas. Por ejemplo, la muestra puede tratarse para lisar las células, utilizando tampones de lisis conocidos, sonicación, electroporación, etc., realizándose la purificación y amplificación como se describe a continuación, según sea necesario, como apreciarán los expertos en la técnica. Además, las reacciones descritas en este documento pueden realizarse de varias maneras, como

apreciarán los expertos en la técnica. Los componentes de la reacción pueden agregarse simultáneamente, o secuencialmente, en cualquier orden, con las formas de realización preferidas que se describen a continuación. Además, la reacción puede incluir una variedad de otros reactivos que pueden incluirse en los ensayos. Estos incluyen reactivos como sales, tampones, proteínas neutras, p. ej. albúmina, detergentes, etc., que pueden usarse para facilitar la hibridación y detección óptimas, y/o reducir las interacciones no específicas o de fondo. También se pueden usar reactivos que, de otro modo, mejoren la eficiencia del ensayo, como inhibidores de proteasa, inhibidores de nucleasa, agentes antimicrobianos, etc., dependiendo de los métodos de preparación de la muestra y la pureza de la diana.

Además, en la mayoría de las realizaciones, los polinucleótidos diana bicatenarios se desnaturalizan para hacerlos monocatenarios a fin de permitir la hibridación de los cebadores y otras sondas descritas en el presente documento.

Una realización utiliza una etapa térmica, generalmente elevando la temperatura de la reacción a aproximadamente 95°C, aunque también se pueden usar cambios de pH y otras técnicas.

Como se describe en este documento, el polinucleótido diana puede ser un producto de una reacción tal como una secuencia de detección de una reacción, una sonda ligada, una sonda extendida de una reacción de PCR o un producto de amplificación de PCR ("amplicón"), etc.

En algunas realizaciones, el polinucleótido diana comprende una posición para la cual se desea información de secuencia, generalmente referida en este documento como la "posición de detección". En una cierta realización, la posición de detección es un solo nucleótido, aunque en algunas realizaciones, puede comprender una pluralidad de nucleótidos, ya sea contiguos entre sí o separados por uno o más nucleótidos. Por "pluralidad", como se usa en este documento, se entienden al menos dos. Como se usa en este documento, la base que se aparea con una base de posición de detección en un híbrido se denomina "posición de lectura" o "posición de interrogación"; por lo tanto, muchas de las sondas de la primera o segunda etapa comprenden una posición de interrogación.

Los métodos descritos en este documento pueden adoptar una amplia variedad de configuraciones, como se muestra en las figuras y se describen con más detalle en este documento. En general, estos componentes incluyen un componente de reducción de complejidad, un componente de especificidad y un componente de amplificación. Los componentes se pueden configurar de varias maneras como se describe a continuación. Es decir, en una realización, primero se realiza una etapa de reducción de complejidad. A esto le sigue la etapa de amplificación o especificidad. Alternativamente, la etapa de especificidad se realiza primero. Esto puede ser seguido por la reducción de complejidad o la etapa de amplificación. Alternativamente, la amplificación se realiza primero. A esto le siguen las etapas de complejidad y especificidad.

Mientras que lo anterior indica que cada uno de los tres componentes se puede realizar en cualquier orden, cualquier experto en la materia apreciará que cuando la amplificación se realiza primero, probablemente habrá algún grado de reducción de complejidad o de la especificidad involucrada. Además, cuando los componentes de especificidad se realicen primero, habrá un grado de reducción de la complejidad. Además, en algunas realizaciones, cuando se realice la amplificación por primera vez, habrá cierto grado de especificidad y reducción de la complejidad. Sin embargo, como se describe a continuación, el método generalmente incluye tres componentes.

Sondas y cebadores

Como aprecia cualquier experto en la técnica, existen varias sondas o cebadores que pueden usarse en los métodos descritos en el presente documento. Estas sondas/cebadores pueden tomar una variedad de configuraciones y pueden tener una variedad de componentes estructurales que se describen con más detalle a continuación. La sonda de la primera etapa puede ser una sonda específica de alelo o una sonda específica de locus. Por sonda o cebador "específico de alelo" se entiende una sonda o cebador que se hibrida con una secuencia diana y discrimina entre alelos o se hibrida con una secuencia diana y se modifica de una manera específica de alelo. Por sonda o cebador "específico del lugar" se entiende una sonda o cebador que se hibrida con una secuencia diana de una manera específica del lugar, pero no necesariamente discrimina entre alelos. Un cebador específico de locus también puede modificarse, es decir, extenderse como se describe a continuación, de manera que incluya información sobre un alelo particular, pero el cebador específico de locus no discrimina entre alelos.

En muchas realizaciones, las sondas o cebadores comprenden uno o más sitios de cebado universales y/o secuencias de identificación. Por ejemplo, en una configuración, cada una de las cuatro bases de alelos está asociada con una secuencia diferente, es decir, la secuencia de identificación del alelo (por ejemplo, el marcador o señuelo), teniendo cada secuencia eficiencias de amplificación similares. En otra configuración, una de las sondas incluye una secuencia de identificación de locus (por ejemplo, el marcador o señuelo).

El tamaño del cebador y el ácido nucleico de la sonda puede variar, como apreciarán los expertos en la materia con cada porción de la sonda y la longitud total de la sonda en general varía de 5 a 500 nucleótidos de longitud. Cada porción puede tener entre 10 y 300, entre 15 y 250, o entre 10 y 35 nucleótidos de longitud, según la técnica de uso y amplificación. Así, por ejemplo, el(los) sitio(s) de cebado universal de las sondas pueden tener una longitud de entre 15 y 20 nucleótidos, utilizándose 18 en ciertas realizaciones. El locus y/o las secuencias de identificación de alelos de las sondas pueden tener una longitud de entre 10 y 300 nucleótidos, utilizándose de 20 a 100 en ciertas realizaciones. La porción específica de la diana de la sonda puede tener una longitud de 15 a 50 nucleótidos. Además, el cebador

puede incluir un sitio de cebado de amplificación adicional.

En una realización, la sonda o sondas específicas de alelo o locus comprenden un dominio diana sustancialmente complementario a un primer dominio de la secuencia diana. En general, las sondas pueden diseñarse para ser complementarias con una secuencia diana (ya sea la secuencia diana de la muestra o a otras secuencias de la sonda, como se describe en el presente documento), de modo que se produzca la hibridación de la diana y las sondas descritas en el presente documento. Esta complementariedad no necesita ser perfecta; puede haber cualquier número de cotejos erróneos de pares de bases que interfieran con la hibridación entre la secuencia diana y los ácidos nucleicos monocatenarios. Sin embargo, si el número de mutaciones es tan grande que no se puede producir una hibridación en las condiciones de hibridación más estrictas, la secuencia no es una secuencia diana complementaria. Así, por "sustancialmente complementario" en el presente documento se entiende que las sondas son suficientemente complementarias con las secuencias diana para hibridar en las condiciones de reacción seleccionadas.

Además, las sondas utilizadas en los métodos descritos en el presente documento pueden construirse de modo que contengan el sitio o sitios de cebado necesarios para el esquema de amplificación posterior. En ciertas realizaciones, los sitios de cebado son sitios de cebado universales. Por "sitio de cebado universal" o "secuencias de cebado universales" en el presente documento se entiende una secuencia de la sonda que se unirá a un cebador para la amplificación.

Como apreciarán los expertos en la técnica, en general, pueden realizarse reacciones altamente multiplexadas, siendo todos los sitios de cebado universales iguales para todas las reacciones. Alternativamente, se pueden usar "conjuntos" de sitios de cebado universales y sondas correspondientes, ya sea simultánea o secuencialmente. Los sitios de cebado universales se usan para amplificar las sondas modificadas para formar una pluralidad de amplicones que luego se detectarán de varias maneras, como se describe en este documento.

Por consiguiente, los métodos descritos en este documento proporcionan primeros conjuntos de sondas diana. Por "conjunto de sondas" en el presente documento se entiende una pluralidad de sondas diana que se utilizan en un ensayo multiplexado particular. En este contexto, pluralidad significa al menos dos, prefiriéndose más de 10, dependiendo del ensayo, de la muestra y el propósito de la prueba. En una realización, el conjunto de sondas incluye más de 100, prefiriéndose más de 500 sondas y siendo particularmente preferidas más de 1000. En una realización particularmente preferida, cada sonda contiene al menos 5000, siendo las más preferidas más de 10.000 sondas.

Componente de reducción de complejidad

La reducción de la complejidad puede ser un componente del esquema múltiplex establecido en este documento. En general, la reducción de complejidad es un método de enriquecimiento para una diana o lugar en particular. Es decir, la reducción de la complejidad se considera un método que da como resultado la eliminación de ácidos nucleicos no diana de la muestra o la eliminación de sondas/cebadores que no se han hibridado correctamente o en absoluto con un ácido nucleico diana. Además, la reducción de la complejidad incluye la eliminación de sondas que no se han modificado durante una etapa enzimática. Es decir, la reducción de complejidad incluye la eliminación de ácidos nucleicos no diana, es decir, el enriquecimiento para ácidos nucleicos diana o la eliminación de sondas o cebadores no hibridados antes de una etapa enzimática, es decir, una etapa de amplificación o especificidad, o ambas.

Hay una variedad de métodos que incluyen una etapa de reducción de la complejidad. Estos incluyen, entre otros, la inmovilización selectiva de ácidos nucleicos o sondas/cebadores diana que se modifican de una manera específica, la eliminación selectiva de ácidos nucleicos no diana y la destrucción selectiva de ácidos nucleicos no diana. Dicha destrucción incluye, entre otros, la desnaturalización, degradación o escisión de ácidos nucleicos no diana. Además, la reducción de la complejidad puede incluir componentes como la amplificación selectiva de la diana, aunque esto también incluye la amplificación y los componentes.

En ciertas realizaciones, la reducción de la complejidad se logra mediante la inmovilización selectiva de un cebador que se ha modificado de una manera específica de la diana. Es decir, los cebadores específicos de locus o específicos de alelo se hibridan con una diana. La diana puede estar inmovilizada o en solución. Después de la hibridación, el cebador se amplifica en una reacción de amplificación del cebador. En algunos aspectos, el cebador o los NTP incluyen un marcador de purificación que permite la eliminación o purificación del producto ampliado de la mezcla de reacción. Una vez amplificado, generalmente el cebador modificado puede inmovilizarse sobre un soporte sólido. Después de la inmovilización del cebador modificado, el soporte se puede lavar para eliminar tanto los ácidos nucleicos no diana como los cebadores que no se modificaron, es decir, que no se amplificaron. Los cebadores inmovilizados, por lo tanto, incluyen información sobre el locus diana, incluida la información alélica particular. Esto da como resultado el enriquecimiento de ácidos nucleicos diana o la eliminación de ácidos nucleicos no diana.

En otra realización, el componente de reducción de complejidad incluye la inmovilización selectiva del polinucleótido diana. Es decir, los polinucleótidos diana se inmovilizan preferentemente sobre un soporte sólido en lugar de ácidos nucleicos no diana.

En una realización, el polinucleótido, sonda o cebador diana, que incluye un cebador modificado, está unido a un soporte sólido. Por "soporte sólido" u otros equivalentes gramaticales en el presente documento se entiende cualquier material que sea apropiado o que pueda modificarse para que sea apropiado para la unión de las secuencias diana.

Como apreciarán los expertos en la técnica, el número de sustratos posibles es muy grande. Los posibles sustratos incluyen, entre otros, vidrio y vidrio modificado o funcionalizado, plásticos (incluidos los acrílicos, poliestireno y copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos, Teflon™, etc.), polisacáridos, nylon o nitrocelulosa, cerámica, resinas, sílice o materiales a base de sílice, incluidos silicio y silicio modificado, carbono, metales, vidrios inorgánicos, plásticos, haces de fibra óptica y una variedad de otros polímeros. Las perlas magnéticas y las placas de microtitulación de alto rendimiento son particularmente preferidas.

La composición y la geometría del soporte sólido varían con su uso. En ciertas realizaciones, se pueden usar soportes que comprenden microesferas o perlas para el soporte sólido. Por "microesferas" o "perlas" o equivalentes gramaticales en el presente documento se entiende pequeñas partículas discretas. La composición de las perlas variará, dependiendo de la clase de agente bioactivo y el método de síntesis. Las composiciones de perlas adecuadas incluyen aquellas utilizadas en la síntesis de péptidos, ácidos nucleicos y restos orgánicos, que incluyen, entre otros, plásticos, cerámicas, vidrio, poliestireno, metilestireno, polímeros acrílicos, materiales paramagnéticos, toria sol, carbono graficado, dióxido de titanio, látex, dextranos reticulados tales como sefarosa, celulosa, nailon, micelas reticuladas y teflón, todos se pueden usar así como cualquier otro material descrito en este documento para los soportes sólidos. La "Microsphere Detection Guide" de Bangs Laboratories, Fishers Ind. es una guía útil. Preferiblemente, en esta realización, cuando se realiza una reducción de complejidad, las microesferas son microesferas o perlas magnéticas.

Una vez unidos al soporte sólido, la secuencia diana, la sonda o los cebadores son susceptibles de análisis como se describe en este documento.

Se puede usar una variedad de condiciones de hibridación o lavado en la presente invención, que incluyen condiciones de rigurosidad alta, moderada y baja; véase, por ejemplo, Maniatis et al., *Molecular Cloning: A Laboratory Manual*, 2ª Edición, 1989, y *Short Protocols in Molecular Biology*, ed. Ausubel, et al. Las condiciones estrictas dependen de la secuencia y serán diferentes en diferentes circunstancias. Las secuencias más largas se hibridan específicamente a temperaturas más altas. Una guía extensa para la hibridación de ácidos nucleicos se encuentra en Tijssen, *Techniques in Biochemistry and Molecular Biology Hybridization with Nucleic Acid Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assay" (1993). En general, las condiciones rigurosas se seleccionan para que sean aproximadamente 5-10°C más bajas que el punto de fusión térmico (Tm) para la secuencia específica a una fuerza iónica y pH definidos. La Tm es la temperatura (bajo la fuerza iónica definida, el pH y la concentración de ácidos nucleicos) a la cual el 50% de las sondas complementarias con la diana se hibridan con la secuencia diana en equilibrio (ya que las secuencias diana están presentes en exceso, a la Tm, 50% de las sondas están ocupadas en el equilibrio). Las condiciones más estrictas serán aquellas en las que la concentración de sal sea inferior a aproximadamente 1,0 M de iones de sodio, típicamente de aproximadamente 0,01 a 1,0 M de concentración de iones de sodio (u otras sales) a pH de 7,0 a 8,3 y la temperatura es de al menos aproximadamente 30°C para sondas cortas (por ejemplo, de 10 a 50 nucleótidos) y al menos aproximadamente 60°C para sondas largas (por ejemplo, más de 50 nucleótidos). También se pueden lograr condiciones rigurosas con la adición de agentes desestabilizantes de hélice, como la formamida.

Por "enzima de extensión" en el presente documento se entiende que es una enzima que extenderá una secuencia mediante la adición de NTP. Como es conocido en la técnica, existe una amplia variedad de enzimas de extensión adecuadas, de las cuales se prefieren las polimerasas (tanto ARN como ADN, dependiendo de la composición de la secuencia diana y la sonda precircular). Las polimerasas preferidas son aquellas que carecen de actividad de desplazamiento de cadena, de modo que sean capaces de agregar solo las bases necesarias al final de la sonda, sin extender más la sonda para incluir nucleótidos que sean complementarios a un dominio de direccionamiento y, por lo tanto, eviten la circularización. Las polimerasas adecuadas incluyen, entre otras, ADN y ARN polimerasas, incluido el fragmento Klenow de la ADN polimerasa I, SEQUENASE 1.0 y SEQUENASE 2.0 (US Biochemical), ADN polimerasa T5, ADN polimerasa Phi29 y diversas ARN polimerasas como Thermus sp., o Q beta replicasa de bacteriófago, también se pueden usar las ARN polimerasas SP6, T3, T4 y T7, entre otras.

Las polimerasas también pueden incluir aquellas que están esencialmente desprovistas de una actividad de exonucleasa de 5' a 3' para asegurar que la sonda no se extienda más allá del extremo 5' de la sonda. Las enzimas ejemplares que carecen de actividad de exonucleasa de 5' a 3' incluyen el fragmento Klenow de la ADN polimerasa y el fragmento Stoffel de la polimerasa DNAPtaq. Por ejemplo, el fragmento de Stoffel de la ADN polimerasa Taq carece de actividad exonucleasa de 5' a 3' debido a manipulaciones genéticas, lo que resulta en una producción de una proteína truncada que carece de los 289 aminoácidos N-terminales. (Véase, por ejemplo, Lawyer et al., *J. Biol. Chem.*, 264: 6427-6437 (1989); y Lawyer et al., *PCR Meth. Appl.*, 2: 275-287 (1993)). Se han generado polimerasas mutantes análogas para polimerasas derivadas de T. maritima, Tsp17, TZ05, Tth y Taf.

Otras polimerasas son aquellas que carecen de una actividad de exonucleasa de 3' a 5', lo que comúnmente se denomina actividad de lectura de prueba, y que elimina las bases que no coinciden en el extremo 3' de un dúplex de plantilla de cebado. Aunque la presencia de actividad de exonucleasa 3' a 5' proporciona una mayor fidelidad en la cadena sintetizada, la actividad de exonucleasa de 3' a 5' encontrada en ADN polimerasas termoestables como Tma (incluidas las formas mutantes de Tma que carecen de actividad exonucleasa de 5' a 3') también degrada el ADN de una sola hebra, como los cebadores utilizados en la PCR, las plantillas de una sola hebra y los productos PCR de una sola hebra. La integridad del extremo 3' de un cebador oligonucleótido utilizado en un proceso de extensión del cebador es crítica, ya que es desde este extremo que comienza la extensión de la cadena naciente. La degradación del extremo

3' conduce a un oligonucleótido acortado que, a su vez, da como resultado una pérdida de especificidad en la reacción de cebado (es decir, cuanto más corto sea el cebador, más probable será que ocurra un cebado espúreo o no específico).

Sin embargo, otras polimerasas son polimerasas termoestables. Una enzima resistente al calor puede incluir cualquier enzima que retenga la mayor parte de su actividad después de una hora a 40°C en condiciones óptimas. Los ejemplos de polimerasas termoestables que carecen de exonucleasa tanto 5' como 3' y exonucleasa 3' a 5' incluyen el fragmento Stoffel de la ADN polimerasa Taq. Esta polimerasa carece de actividad de exonucleasa 5' a 3' debido a la manipulación genética y no presenta ninguna actividad de 3' a 5', ya que la polimerasa Taq carece naturalmente de actividad de exonucleasa 3' a 5'. La ADN polimerasa Tth se deriva de *Thermus thermophilus* y está disponible en Epicenter Technologies, Molecular Biology Resource Inc. o Perkin-Elmer Corp. Otras ADN polimerasas útiles que carecen de actividad de exonucleasa 3' incluyen un Vent[R](exo-), disponible de New England Biolabs, Inc.0 (purificado a partir de cepas de *E. coli* que portan un gen de ADN polimerasa de la arqueobacteria *Thermococcus litoralis*) y la ADN polimerasa Hot Tub derivada de *Thermus flavus* y disponible de Amersham Corporation. Otras enzimas preferidas que son termoestables y están privadas de actividad de exonucleasa de 5' a 3' y de la actividad de exonucleasa de 3' a 5' incluyen AmpliTaq Gold. Otras ADN polimerasas, que son al menos sustancialmente equivalentes, pueden usarse como otras ADN polimerasas I de *Thermus aquaticus* (Taq) truncadas en el extremo N, la polimerasa denominada KlenTaq I y KlenTaq LA son bastante adecuadas para ese propósito. Por supuesto, cualquier otra polimerasa que tenga estas características también se puede usar de acuerdo con la invención.

Las condiciones para realizar la adición de uno o más nucleótidos en el extremo 3' de la sonda dependerán de la enzima particular utilizada, y generalmente seguirán las condiciones recomendadas por el fabricante de las enzimas utilizadas.

Componente de especificidad

Generalmente, después de una etapa de reducción de complejidad, se incluye una etapa de especificidad en el método descrito en este documento. Por "componente de especificidad" se entiende una etapa que discrimina entre ácidos nucleicos diana, preferiblemente a nivel del alelo. Es decir, el componente de especificidad es una etapa específica del alelo (por ejemplo, genotipado o análisis de SNP). Si bien puede lograrse cierto nivel de especificidad mediante la simple hibridación de sondas específicas de alelo con la plantilla (es decir, el producto de la etapa de reducción de complejidad anterior), en una realización preferida, la etapa de especificidad incluye una etapa enzimática. Es decir, la fidelidad de una etapa enzimática mejora la especificidad para la discriminación de alelos. Las enzimas preferidas incluyen ADN polimerasas, ARN polimerasas y ligasas como se describe con más detalle en el presente documento.

Las polimerasas como se describieron anteriormente también pueden ser adecuadas para las etapas de especificidad.

Se conocen muchas ligasas y son adecuadas para su uso en los métodos descritos en el presente documento. Las ligasas ejemplares se describen en Lehman, Science, 186: 790-797 (1974); Engler et al, DNA Ligases, páginas 3-30 en Boyer, editor, The Enzymes, vol. 15B (Academic Press, Nueva York, 1982); y similares. Las ligasas preferidas incluyen ADN ligasa T4, ADN ligasa T7, ADN ligasa de *E. coli*, ligasa Taq, ligasa Pfu y ligasa Tth. Los protocolos para su uso son conocidos, por ejemplo, en Sambrook et al. (citado anteriormente); Barany, PCR Methods and Applications, 1: 5-16 (1991); Marsh et al., Strategies, 5: 73-76 (1992); y similares. En general, las ligasas requieren que esté presente un grupo fosfato 5' para la ligadura al hidroxilo 3' de una cadena contigua. Las ligasas preferidas incluyen ligasas termoestables o termófilas, tales como pfu ligasa, ligasa Tth, ligasa Taq y ADN ligasa Ampligase™ (Epicenter Technologies, Madison, WI). La Ampligase tiene una baja actividad de ligadura del extremo romo.

En ciertas realizaciones, la ligasa es una que tiene la menor unión de apareamiento erróneo. La especificidad de la ligasa puede aumentarse sustituyendo las ligasas dependientes de NAD⁺ más específicas, como la ligasa de *E. coli* y la ligasa Taq (termoestable) por la ADN ligasa T4 menos específica. El uso de análogos de NAD en la reacción de unión aumenta aún más la especificidad de la reacción de unión. Ver la Patente de EE.UU. N° 5.508.179 de Wallace et al.

En una realización, el componente de especificidad se realiza con dianas inmovilizadas. Es decir, los productos de la etapa de reducción de complejidad se inmovilizan en un soporte sólido como se describe en este documento. Como se analiza en el presente documento, la diana de la reacción de especificidad se conoce como "diana de especificidad". Es decir, el producto de la etapa de reducción de complejidad es la diana de especificidad.

En una realización, el soporte es el mismo que en la etapa inicial de reducción de complejidad. En esta realización, el ácido nucleico diana se elimina del soporte sólido antes del ensayo de especificidad. El ácido nucleico diana puede eliminarse mediante cualquier método que desnaturalice el complejo de hibridación, dando como resultado la liberación del ácido nucleico diana. Como puede apreciar cualquier experto en la materia, en esta realización, el ácido nucleico diana no está unido covalentemente al soporte sólido. Es decir, es la sonda diana la que está unida de manera estable al soporte. Es decir, aunque la unión de la sonda no es necesariamente covalente, es lo suficientemente estable como para resistir la desnaturalización del complejo de hibridación y la eliminación del ácido nucleico diana no unido.

En una realización alternativa, la diana de especificidad está en solución. Es decir, después de una etapa de reducción

de la complejidad, el complejo de hibridación entre el ácido nucleico diana inmovilizado y la sonda diana se desnaturaliza y la sonda diana modificada se eluye del complejo de hibridación. En una cierta realización, la diana de especificidad se analiza en solución. En una realización alternativa, la diana de especificidad de la fase de solución se inmoviliza en un soporte sólido posterior.

- 5 Estos ensayos de especificidad, es decir, las técnicas de genotipado, se dividen en cinco categorías generales: (1) técnicas que se basan en métodos de hibridación tradicionales que utilizan la variación de las condiciones de rigurosidad (temperatura, condiciones del tampón, etc.) para distinguir los nucleótidos en la posición de detección; (2) técnicas de extensión que agregan una base ("la base") para unirse con el nucleótido en la posición de detección; (3) técnicas de unión, que dependen de la especificidad de las enzimas ligasa (o, en algunos casos, de la especificidad de las técnicas químicas), de modo que las reacciones de unión se producen preferentemente si existe una complementariedad perfecta en la posición de detección; (4) técnicas de escisión, que también se basan en la especificidad enzimática o química, de modo que la escisión se produce preferentemente si existe una complementariedad perfecta; y (5) técnicas que combinan estos métodos. Ver, en general, las patentes de EE.UU. Nos. 6.890.741, 6.913.884, 7.955.794, 7.582.420 y 8.288.103, y la Publicación de Estados Unidos 2013-0244882.
- 10 En ciertas realizaciones, se realiza genotipificación de extensión. En esta realización, se puede usar cualquier número de técnicas para agregar un nucleótido a la posición de lectura de una sonda hibridada a la secuencia diana adyacente a la posición de detección. Al basarse en la especificidad enzimática, se agrega preferentemente una base perfectamente complementaria. Algunos de los métodos descritos en el presente documento se basan en la incorporación enzimática de nucleótidos en la posición de detección. Esto se puede hacer utilizando cualquier número de métodos conocidos en la técnica, tales como la extensión de base única o la extensión de múltiples bases. En ciertas realizaciones, el genotipado se realiza mediante la extensión del cebador que no utiliza nucleótidos que terminan la cadena. Como tal, este genotipado se considera extensión multibase. El método incluye proporcionar un oligonucleótido problema diseñado para detectar un alelo de un SNP dado. El número de oligonucleótidos se determina por el número de alelos SNP distintos que se están sondeando. Por ejemplo, si se estuviera sondeando 1000 SNP, cada uno con dos alelos, serían necesarios 2000 oligonucleótidos. Los problemas son complementarios a un tramo de ADN que contiene el SNP, con la base terminal de cada problema correspondiente a la posición del SNP, o con la posición específica del SNP dentro de los últimos 1, 2, 3 ó 4 nucleótidos del problema. En algunas realizaciones, el problema no es la posición terminal del cebador, sino que se encuentra en una posición 1, 2, 3, 4, 5 ó 6 nucleótidos del extremo 3' del cebador. Por ejemplo, cuando un SNP tiene un alelo A y C, se proporcionan problemas que terminan en T y G y, en algunas realizaciones, se pueden inmovilizar en elementos separados (perlas) para detectar los dos. Aunque tanto el apareamiento como el apareamiento erróneo se hibridarán con un alelo dado, solo el apareamiento puede actuar como un cebador para una reacción de amplificación de la ADN polimerasa. Por consiguiente, después de la hibridación de las sondas con el ADN diana, se realiza una reacción de polimerasa. Esto da como resultado la amplificación de los híbridos con una ADN polimerasa en presencia de dNTP.
- 20 En ciertas realizaciones, es deseable eliminar las sondas o cebadores no amplificados o sin reaccionar de la mezcla de ensayo y, particularmente, de un soporte sólido, ya que las sondas o cebadores no amplificados pueden competir con los cebadores amplificados en la unión a las sondas de captura. La concentración de los cebadores no amplificados en relación con el cebador amplificado puede ser relativamente alta, ya que generalmente se requiere un gran exceso de cebador para generar una hibridación eficaz del cebador. Por consiguiente, se pueden usar varias técnicas diferentes para facilitar la eliminación de sondas o cebadores no amplificados. Estos generalmente incluyen métodos basados en la eliminación de cebadores sin reaccionar mediante la unión a un soporte sólido, protegiendo los cebadores reaccionados y degradando los no amplificados, y separando los cebadores sin reaccionar y reaccionados.

Componente de amplificación

- 45 En esta realización, en el presente documento se proporcionan métodos que incluyen la amplificación de un polinucleótido y pueden usarse en los métodos para caracterizar un polinucleótido los productos de las reacciones de amplificación de ácidos nucleicos, es decir, los amplicones. Los métodos de amplificación adecuados incluyen tanto la amplificación de la diana como la amplificación de la señal. La amplificación de la diana implica la amplificación (es decir, la replicación) de la secuencia diana a detectar, lo que resulta en un aumento significativo en el número de moléculas diana. Las estrategias de amplificación de la diana incluyen, entre otras, la reacción en cadena de la polimerasa (PCR), la amplificación por desplazamiento de cadena (SDA), la amplificación basada en la secuencia de ácido nucleico (NASBA) y la amplificación de círculo rodante (RCA). Dichas estrategias de amplificación son conocidas por los expertos en la técnica y pueden seleccionarse fácilmente para su uso en los métodos descritos.
- 50 Alternativamente, en lugar de amplificar la diana, otras técnicas usan la diana como una plantilla para replicar una sonda de señalización, lo que permite que un pequeño número de moléculas dianas produzcan un gran número de sondas de señalización, que luego pueden detectarse. Las estrategias de amplificación de señal incluyen la reacción en cadena de la ligasa (LCR), la tecnología de sonda de ciclación (CPT), las técnicas de división invasiva como la tecnología Invader™, la tecnología Q-Beta replicasa (QR) y el uso de "sondas de amplificación" como el "ADN ramificado" que da como resultado múltiples sondas de marcadores unidas a una única secuencia diana.
- 55 Todos estos métodos pueden incluir un ácido nucleico cebador (incluidos los análogos de ácido nucleico) que se
- 60

hibrida con una secuencia diana para formar un complejo de hibridación, y se agrega una enzima que, de alguna manera, modifica el cebador para formar un cebador modificado. Por ejemplo, la PCR generalmente requiere dos cebadores, dNTP y una ADN polimerasa; la LCR requiere dos cebadores que hibridan adyacentemente a la secuencia diana y una ligasa; CPT requiere un cebador escindible y una enzima de escisión; la escisión invasiva requiere dos cebadores y una enzima de escisión; etc. Así, en general, se agrega un ácido nucleico diana a una mezcla de reacción que comprende los componentes de amplificación necesarios, y se forma un cebador modificado.

En general, el cebador modificado sirve como secuencia diana para una reacción secundaria, que luego produce una serie de cadenas amplificadas, que se pueden detectar como se describe en este documento. Según se requiera, los cebadores sin reaccionar se eliminan, de diversas maneras, como apreciarán los expertos en la técnica y se describen en este documento. Por consiguiente, la reacción comienza con la adición de un ácido nucleico cebador a la secuencia diana que forma un complejo de hibridación. Una vez que se ha formado el complejo de hibridación entre el cebador y la secuencia diana, se utiliza una enzima, a veces denominada "enzima de amplificación", para modificar el cebador. En cuanto a todos los métodos descritos en este documento, las enzimas se pueden agregar en cualquier punto durante el ensayo, ya sea antes, durante o después de la adición de los cebadores. La identidad de la enzima dependerá de la técnica de amplificación utilizada. Del mismo modo, la modificación dependerá de la técnica de amplificación.

En ciertas realizaciones, la técnica de amplificación de la diana es la reacción en cadena de la polimerasa (PCR). La PCR se usa y describe ampliamente, e implica el uso de la amplificación del cebador combinado con el ciclo térmico para amplificar una secuencia diana; ver las patentes de EE.UU. Nos. 4.683.195 y 4.683.202, y PCR Essential Data, J. W. Wiley & sons, Ed. C. R. Newton, 1995. Además, hay una serie de variaciones de la PCR que también encuentran uso, incluyendo "PCR cuantitativa competitiva" o "QC-PCR", "PCR con cebado arbitrario" o "AP-PCR", "inmuno-PCR", "Alu-PCR", "PCR polimorfismo conformacional de una sola hebra" o "PCR-SSCP", "PCR con transcriptasa inversa" o "RT-PCR", "PCR de captura de biotina", "PCR de vectorette", "PCR de manillar" y "PCR con selección de resta de ADNc", "PCR específica de alelo", entre otras. Se entiende que cualquier persona con experiencia ordinaria en la técnica podrá seleccionar fácilmente la variación apropiada de la PCR que podría usarse en los métodos descritos en el presente documento.

En ciertas realizaciones, la reacción de amplificación es una reacción de amplificación múltiple tal como se describe en el presente documento. En una realización, la reacción de amplificación utiliza una pluralidad de cebadores de PCR para amplificar una pluralidad de secuencias diana. En esta realización, la pluralidad de secuencias diana se amplifica simultáneamente con la pluralidad de pares de cebadores de amplificación.

Una realización alternativa de la reacción de PCR múltiple utiliza cebadores universales como se describe en este documento. Es decir, los cebadores de la PCR universales se hibridan con los sitios de cebado universales en la secuencia diana y, por lo tanto, amplifican una pluralidad de secuencias diana. Esta realización es potencialmente preferida porque requiere solo un número limitado de cebadores de PCR. Es decir, tan solo un par de cebadores puede amplificar una pluralidad de secuencias diana.

Los amplicones Golden Gate se generaron utilizando ADN humano como plantilla como se describió anteriormente (Cold Spring Harb Symp Quant Biol. 2003; 68:69-78. Highly parallel SNP genotyping. Fan J B et al.). Los amplicones resultantes tenían uno de los dos cebadores, designados P1 y P2, dependiendo del alelo. Además, el cebador inverso universal ("Reverse P3") estaba presente en todos los amplicones.

P1: TCTCGTCGCTCATCAACT (SEC ID NO: 82)

P2: GAGTCGAGGTCATATCGT (SEC ID NO: 83)

P3 Inversa: GTCTGCCTATAGTGAGTC (SEC ID NO: 84)

Se utilizó una segunda ronda de PCR que empleaba 16 ciclos para agregar los cebadores de señuelos del alelo, denominados "P1_barcode_A" y "P1_barcode_B". Se usó un cebador inverso universal amplificado ("Universal dU Reverse") que contenía múltiples residuos de desoxiuracilo.

P1_barcode_A:

/5phos/TTTTTTTTTTTTTTTCCTTCCXXTTTTCTTCTTCTTCAAGAAGAAGATCTCGT
CGCTCATCAACT (SEQ ID NO: 85)

P2_barcode_B:

/5phos/TTTTTTTTTTTTTTTCCTTCCXXTTTTTAATTAATTTTGTGTTGTGAGTCG
AGGTCATATCGT (SEQ ID NO: 86)

Universal dU Reverse:

ATACGGCG/dU/CCACCGACC/dU/CAGCGTC/dU/GCCTATAG/dU/GAGTC (SEQ ID

NO: 87)

Donde /5phos/ indica un fosfato 5', /dU/ es una base de desoxiuracilo, y X es un resto abásico.

- 5 Después de la PCR, la muestra se incubó con la enzima USER (New England Biolabs, Ipswich, MA) durante 2,5 horas a 37°C para crear espacios de una sola hebra donde se localizó un residuo dU. La muestra se calentó a 65°C durante 10 min para eliminar el ADN fragmentado y crear un saliente de 3'. La muestra se purificó utilizando un kit de limpieza de PCR (Qiagen).

Las muestras se recocieron en oligo que contenía colesterol "P3_Chol" en una relación molar de 1:1 calentando a 65°C y enfriando lentamente.

- 10 P3_Chol:

ACCGACACTGCGTCTGCCTATAGTGAGTC/iSp9//3CholTEG/ (SEC ID NO: 88)

Donde /iSp9/ indica un espaciador de trietilenglicol de 9 átomos, y /3CholTEG/ indica un resto TEG (trietilenglicol) de colesterol 3'.

- 15 Las bicapas lipídicas se formaron a partir de 1,2-difitoil-sn-glicero-3-fosfocolina (lípidos polares de Avanti). La bicapa cubría una abertura horizontal de ~20 micras de diámetro en teflón. Se agregó M2-NNN-MspA al lado conectado a tierra de la bicapa a una concentración de ~2,5 ng/ml. Una vez que se insertó un solo poro, el compartimento se lavó con un tampón experimental para evitar futuras inserciones. Un amplificador de pinza de parche Axopatch-200B (Axon Instruments) aplicó un voltaje a través de la bicapa de 180 mV y midió las corrientes iónicas. La señal analógica se filtró a paso bajo a 50 kHz con un filtro Bessel de 4 polos y luego se digitalizó a cinco veces la frecuencia del filtro a
- 20 paso bajo. La adquisición de datos se controló con un software personalizado escrito en LabWindows/CVI (National Instruments).

Los compartimentos de ~60 µl en ambos lados de la bicapa contenían un tampón experimental de KCl 0,4 M, EDTA 1 mM, DTT 1 mM, ATP 1 mM, MgCl₂ 10 mM y HEPES/KOH 10 mM tamponados a pH 8,0. Se usó Tga natural Hel308 como motor a 150 nM.

- 25 La Fig. 23 ilustra señales simuladas ejemplares que pueden generarse en función del tiempo para una primera secuencia polinucleotídica ilustrativa (SEC ID NO: 89) y una segunda secuencia polinucleotídica ilustrativa (SEC ID NO: 90) adecuada para su uso como señuelos respectivos, según algunas realizaciones. Puede verse en la fig. 24 que la señal simulada (1) correspondiente a la translocación fraccional de la primera secuencia polinucleotídica
- 30 ilustrativa a través de un poro por una helicasa Hel308 tiene un patrón característico en el tiempo que incluye dos "picos" a un nivel de señal relativamente alto, seguido de una caída, seguido por dos "picos" más a un nivel de señal relativamente bajo, mientras que la señal simulada (2) correspondiente a la translocación fraccional de la segunda secuencia polinucleotídica ilustrativa a través de un poro por una helicasa Hel308 incluye dos "picos" a un nivel de señal relativamente bajo, seguido por un aumento, seguido de dos "picos" más a un nivel de señal relativamente alto.
- 35 Por consiguiente, puede esperarse que las señales reales que incluyen características distintivas, como en las señales simuladas (1) y (2), puedan distinguirse fácilmente entre sí, por ejemplo, utilizando la coincidencia de patrones y, por lo tanto, pueden facilitar distinguir los resultados de los ensayos entre sí. Por ejemplo, las figs. 24A-24D ilustran señales simuladas ejemplares que pueden generarse en función del tiempo para la primera y segunda secuencias polinucleotídicas ilustrativas adecuadas para su uso como señuelos respectivos, de acuerdo con algunas realizaciones. Se puede ver que las secciones de las señales simuladas generalmente designadas "señuelo" en las
- 40 Figs. 24A y 24B incluyen dos "picos" a un nivel de señal relativamente alto, seguidos de una caída, seguidos por otros dos "picos" a un nivel de señal relativamente bajo y, por lo tanto, se puede entender que corresponden a la primera secuencia polinucleotídica ilustrativa. También se puede ver que las secciones de las señales simuladas generalmente designadas "señuelo" en las Figs. 24C y 24D incluyen dos "picos" a un nivel de señal relativamente bajo, seguidos de un aumento, seguidos por dos "picos" más a un nivel de señal relativamente alto, y, por lo tanto, puede entenderse que corresponden a la segunda secuencia polinucleotídica ilustrativa.
- 45

- En otro ejemplo, se insertó un poro de 2NNN MspA en una bicapa lipídica DPhPC de una manera análoga a la descrita anteriormente en el presente ejemplo. El tampón incluía KCl 400 mM, HEPES 10 mM, pH 8, MgCl₂ 5 mM y EDTA 1 mM. Los reactivos incluían DTT 1 mM y ATP 1 mM. La enzima incluía aproximadamente 150 mM de Hel308 Tga. El ADN fue de aproximadamente 10 nM, y las cadenas individuales que se secuenciaron (indicadas como RS1801131
- 50 SNP1 y SNP2) se hibridaron a un polinucleótido que contenía colesterol. Las señales obtenidas durante la secuenciación de tales cadenas se descifraron utilizando un procesamiento posterior, que incluía la búsqueda de niveles y la alineación con secuencias anticipadas utilizando algoritmos como los descritos en otras partes en este documento.

Las figuras 25A y 25B ilustran respectivamente señales simuladas ejemplares que pueden generarse en función del

tiempo para la primera y segunda secuencias polinucleotídicas ilustrativas adecuadas para su uso como señuelos respectivos, de acuerdo con algunas realizaciones. Las secciones de las señales simuladas dentro de los cuadros de puntos en las Figs. 25A y 25B incluyen respectivamente patrones distintivos para las secuencias designadas respectivamente rs1801131 SNP1 y rs1801131 SNP2 y, por lo tanto, pueden usarse como señuelos respectivos. Las secuencias utilizadas como señuelos fueron las mismas que se ilustran en la FIG. 23.

Las figuras 26A-26D ilustran respectivamente señales medidas ilustrativas que se generaron en función del tiempo para la primera y la segunda secuencia polinucleotídica ilustrativa adecuadas para su uso como señuelos respectivos, de acuerdo con algunas realizaciones. Las secciones de las señales medidas dentro de los cuadros de puntos en las Figs. 26A y 26B incluyen respectivamente patrones distintivos que pueden verse que corresponden al señuelo de la secuencia designada rs1801131 SNP1, mientras que las secciones de las señales medidas dentro de los cuadros de puntos en las Figs. 26C y 26D incluyen respectivamente patrones distintivos que pueden verse que corresponden al señuelo de la secuencia designada rs1801131 SNP2, y también se pueden distinguir fácilmente del señuelo de la secuencia designada rs1801131 SNP1.

Otras realizaciones alternativas

Se debe tener en cuenta que los sistemas y métodos proporcionados en el presente documento pueden implementarse utilizando diversos tipos de entornos de procesadores de datos (por ejemplo, en uno o más procesadores de datos) que ejecutan instrucciones (por ejemplo, instrucciones de software) para realizar las operaciones descritas en este documento. Los ejemplos no limitantes incluyen la implementación en una sola computadora o dispositivo de uso general, en un sistema en red, en una configuración cliente-servidor o en una configuración de proveedor de servicios de aplicaciones. Por ejemplo, los métodos y sistemas descritos en este documento pueden implementarse en muchos tipos diferentes de dispositivos de procesamiento mediante un código de programa que comprenda instrucciones de programa que sean ejecutables por el subsistema de procesamiento del dispositivo. Las instrucciones del programa de software pueden incluir un código fuente, código objeto, código máquina o cualquier otro dato almacenado que sea operable para hacer que el sistema de procesamiento realice los métodos y operaciones descritas en este documento. Sin embargo, también se pueden usar otras implementaciones, como el firmware o incluso el hardware diseñado apropiadamente configurado para llevar a cabo los métodos y sistemas descritos en este documento.

También se señala que los sistemas y métodos pueden incluir señales de datos transmitidas a través de redes (por ejemplo, red de área local, red de área amplia, internet, combinaciones de las mismas, etc.), medio de fibra óptica, ondas portadoras, redes inalámbricas, etc. para la comunicación con uno o más dispositivos de procesamiento de datos. Las señales de datos pueden transportar cualquiera o todos los datos descritos en este documento que se proporcionan al dispositivo o desde el dispositivo.

Los datos de los sistemas y los métodos (por ejemplo, asociaciones, entrada de datos, salida de datos, resultados de datos intermedios, resultados de datos finales, etc.) se pueden almacenar e implementar en uno o más tipos diferentes de almacenes de datos implementados en computadoras, como diferentes tipos de dispositivos de almacenamiento y construcciones de programación (p. ej., RAM, ROM, memoria Flash, archivos planos, bases de datos, estructuras de datos de programación, variables de programación, construcciones de sentencias IF-THEN (o de tipo similar, etc.)). Se observa que las estructuras de datos describen formatos para su uso en la organización y almacenamiento de datos en bases de datos, programas, memoria u otros medios legibles por computadora para el uso de un programa informático.

Otros sistemas y métodos se pueden proporcionar en muchos tipos diferentes de medios de almacenamiento legibles por computadora, incluidos mecanismos de almacenamiento informático (por ejemplo, medios no transitorios, como CD-ROM, disquetes, RAM, memoria flash, discos duros de la computadora, etc.) que contienen instrucciones (por ejemplo, softwares) para su uso en la ejecución por un procesador que realiza las operaciones de los métodos e implemente los sistemas descritos en este documento.

Además, los componentes de la computadora, los módulos de software, las funciones, los almacenamientos de datos y las estructuras de datos proporcionados en este documento pueden conectarse directa o indirectamente entre sí para permitir el flujo de datos necesarios para sus operaciones. También se señala que un módulo o procesador puede incluir, entre otros, una unidad de código que realice una operación de software, y puede implementarse, por ejemplo, como una unidad de código de subrutina, o como una unidad de código de función de software, o como una unidad de código de software, o como un objeto (como en un paradigma orientado a objetos), o como un applet, o en un lenguaje de script de computadora, o como otro tipo de código de computadoras. Los componentes del software y/o la funcionalidad pueden ubicarse en una sola computadora o distribuirse en múltiples computadoras, dependiendo de la situación concreta.

Listado de secuencias

<110> ILLUMINA, INC.

<120> COMPOSITIONES Y MÉTODOS PARA SECUENCIAR NUCLEÓTIDOS

<130> 12957-139-228

<140>
 <141>
 <150> 61/909,316
 <151> 26-11-2013
 5 <160> 90
 <170> PatentIn versión 3.5
 <210> 1
 <211> 184
 <212> PRT
 10 <213> Secuencia Artificial
 <220>
 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: polipéptido sintético"
 <400> 1
 Gly Leu Asp Asn Glu Leu Ser Leu Val Asp Gly Gln Asp Arg Thr Leu
 1 5 10 15
 Thr Val Gln Gln Trp Asp Thr Phe Leu Asn Gly Val Phe Pro Leu Asp
 20 25 30
 Arg Asn Arg Leu Thr Arg Glu Trp Phe His Ser Gly Arg Ala Lys Tyr
 35 40 45
 Ile Val Ala Gly Pro Gly Ala Asp Glu Phe Glu Gly Thr Leu Glu Leu
 50 55 60
 Gly Tyr Gln Ile Gly Phe Pro Trp Ser Leu Gly Val Gly Ile Asn Phe
 65 70 75 80
 Ser Tyr Thr Thr Pro Asn Ile Leu Ile Asn Asn Gly Asn Ile Thr Ala
 85 90 95
 Pro Pro Phe Gly Leu Asn Ser Val Ile Thr Pro Asn Leu Phe Pro Gly
 100 105 110
 Val Ser Ile Ser Ala Arg Leu Gly Asn Gly Pro Gly Ile Gln Glu Val
 115 120 125
 15 Ala Thr Phe Ser Val Arg Val Ser Gly Ala Lys Gly Gly Val Ala Val
 130 135 140
 Ser Asn Ala His Gly Thr Val Thr Gly Ala Ala Gly Gly Val Leu Leu
 145 150 155 160
 Arg Pro Phe Ala Arg Leu Ile Ala Ser Thr Gly Asp Ser Val Thr Thr
 165 170 175
 Tyr Gly Glu Pro Trp Asn Met Asn
 180
 <210> 2
 <211> 4
 20 <212> PRT
 <213> Desconocido

<220>
 <221> fuente
 <223> /nota="Descripción de Desconocido: 'Secuencia de resto de DEAD'-box"

 <400> 2
 Asp Glu Ala Asp
 5 1

 <210> 3
 <211> 4
 <212> PRT
 <213> Desconocido

 10 <220>
 <221> fuente
 <223> /nota="Descripción de Desconocido: 'Secuencia de resto de DEAH'-box"

 <400> 3
 Asp Glu Ala His
 1

 15 <210> 4
 <211> 8
 <212> PRT
 <213> Methanococcoides burtonii

 <400> 4
 Gln Met Ala Gly Arg Ala Gly Arg
 20 1 5

 <210> 5
 <211> 9
 <212> PRT
 <213> Methanococcoides burtonii

 25 <400> 5
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5

 <210> 6
 <211> 8
 <212> PRT
 30 <213> Pyrococcus furiosus

 <400> 6
 Gln Met Leu Gly Arg Ala Gly Arg
 1 5

 <210> 7
 <211> 9
 35 <212> PRT
 <213> Pyrococcus furiosus

 <400> 7
 Gln Met Leu Gly Arg Ala Gly Arg Pro
 1 5

 <210> 8
 <211> 8
 40 <212> PRT
 <213> Haloferax volcanii

 <400> 8
 Gln Met Met Gly Arg Ala Gly Arg
 1 5

 45 <210> 9
 <211> 9
 <212> PRT

<213> Haloferax volcanii

<400> 9
 Gln Met Met Gly Arg Ala Gly Arg Pro
 1 5

5 <210> 10
 <211> 8
 <212> PRT
 <213> Halorubrum lacusprofundi

<400> 10
 Gln Met Cys Gly Arg Ala Gly Arg
 1 5

10 <210> 11
 <211> 8
 <212> PRT
 <213> Halorubrum lacusprofundi

15 <400> 11
 Gln Met Gly Arg Ala Gly Arg Pro
 1 5

<210> 12
 <211> 8
 <212> PRT
 <213> Cenarchaeum symbiosum

20 <400> 12
 Gln Leu Cys Gly Arg Ala Gly Arg
 1 5

<210> 13
 <211> 9
 <212> PRT

25 <213> Cenarchaeum symbiosum

<400> 13
 Gln Leu Cys Gly Arg Ala Gly Arg Pro
 1 5

<210> 14
 <211> 8
 <212> PRT

30 <213> Sulfolobus solfataricus

<400> 14
 Gln Met Ser Gly Arg Ala Gly Arg
 1 5

<210> 15
 <211> 9
 <212> PRT

35 <213> Sulfolobus solfataricus

<400> 15
 Gln Met Ser Gly Arg Ala Gly Arg Pro
 1 5

40 <210> 16
 <211> 8
 <212> PRT
 <213> Methanogenium frigidum

<400> 16
 Gln Met Ala Gly Arg Ala Gly Arg
 1 5

45 <210> 17

<211> 9
 <212> PRT
 <213> Methanogenium frigidum

 <400> 17
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 5 1 5

 <210> 18
 <211> 8
 <212> PRT
 <213> Methanothermococcus okinawensis

 10 <400> 18
 Gln Cys Ile Gly Arg Ala Gly Arg
 1 5

 <210> 19
 <211> 9
 <212> PRT
 15 <213> Methanothermococcus okinawensis

 <400> 19
 Gln Cys Ile Gly Arg Ala Gly Arg Pro
 1 5

 <210> 20
 <211> 8
 20 <212> PRT
 <213> Methanotorris igneus

 <400> 20
 Gln Cys Ile Gly Arg Ala Gly Arg
 1 5

 <210> 21
 <211> 9
 <212> PRT
 25 <213> Methanotorris igneus

 <400> 21
 Gln Cys Ile Gly Arg Ala Gly Arg Pro
 1 5

 30 <210> 22
 <211> 8
 <212> PRT
 <213> Thermococcus gammatolerans

 <400> 22
 Gln Met Met Gly Arg Ala Gly Arg
 35 1 5

 <210> 23
 <211> 9
 <212> PRT
 <213> Thermococcus gammatolerans

 40 <400> 23
 Gln Met Met Gly Arg Ala Gly Arg Pro
 1 5

 <210> 24
 <211> 8
 <212> PRT
 45 <213> Thermococcus barophilus

 <400> 24
 Gln Met Ile Gly Arg Ala Gly Arg
 1 5

	<210> 25
	<211> 9
	<212> PRT
	<213> Thermococcus barophilus
5	<400> 25
	Gln Met Ile Gly Arg Ala Gly Arg Pro
	1 5
	<210> 26
	<211> 8
	<212> PRT
10	<213> Thermococcus sibiricus
	<400> 26
	Gln Met Met Gly Arg Ala Gly Arg
	1 5
	<210> 27
	<211> 9
15	<212> PRT
	<213> Thermococcus sibiricus
	<400> 27
	Gln Met Met Gly Arg Ala Gly Arg Pro
	1 5
20	<210> 28
	<211> 8
	<212> PRT
	<213> Methanosarcina barkeri
	<400> 28
	Gln Met Ala Gly Arg Ala Gly Arg
	1 5
25	<210> 29
	<211> 9
	<212> PRT
	<213> Methanosarcina barkeri
	<400> 29
30	Gln Met Ala Gly Arg Ala Gly Arg Pro
	1 5
	<210> 30
	<211> 8
	<212> PRT
	<213> Methanosarcina acetivorans
35	<400> 30
	Gln Met Ala Gly Arg Ala Gly Arg
	1 5
	<210> 31
	<211> 9
	<212> PRT
40	<213> Methanosarcina acetivorans
	<400> 31
	Gln Met Ala Gly Arg Ala Gly Arg Pro
	1 5
	<210> 32
	<211> 8
45	<212> PRT
	<213> Methanohalophilus mahii
	<400> 32

Gln Met Ala Gly Arg Ala Gly Arg
 1 5
 <210> 33
 <211> 9
 <212> PRT
 5 <213> Methanohalophilus mahii
 <400> 33
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5
 <210> 34
 <211> 8
 10 <212> PRT
 <213> Methanosarcina mazei
 <400> 34
 Gln Met Ala Gly Arg Ala Gly Arg
 1 5
 <210> 35
 <211> 9
 15 <212> PRT
 <213> Methanosarcina mazei
 <400> 35
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5
 20 <210> 36
 <211> 8
 <212> PRT
 <213> Methanosaeta thermophila
 <400> 36
 Gln Met Ala Gly Arg Ala Gly Arg
 25 1 5
 <210> 37
 <211> 9
 <212> PRT
 <213> Methanosaeta thermophila
 30 <400> 37
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5
 <210> 38
 <211> 8
 <212> PRT
 35 <213> Methanosalsum zhilinae
 <400> 38
 Gln Met Ala Gly Arg Ala Gly Arg
 1 5
 <210> 39
 <211> 9
 40 <212> PRT
 <213> Methanosalsum zhilinae
 <400> 39
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5
 <210> 40
 <211> 8
 45 <212> PRT
 <213> Methanohalobium evestigatum

<400> 40
 Gln Met Ala Gly Arg Ala Gly Arg
 1 5

 <210> 41
 <211> 9
 5 <212> PRT
 <213> Methanohalobium evestigatum

 <400> 41
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5

 <210> 42
 10 <211> 8
 <212> PRT
 <213> Methanococcus maripaludis

 <400> 42
 Gln Cys Ile Gly Arg Ala Gly Arg
 1 5

 <210> 43
 15 <211> 9
 <212> PRT
 <213> Methanococcus maripaludis

 <400> 43
 20 Gln Cys Ile Gly Arg Ala Gly Arg Pro

 1 5

 <210> 44
 <211> 8
 25 <212> PRT
 <213> Natrionalba magadii

 <400> 44
 Gln Met Met Gly Arg Ala Gly Arg
 1 5

 <210> 45
 30 <211> 9
 <212> PRT
 <213> Natrionalba magadii

 <400> 45
 Gln Met Met Gly Arg Ala Gly Arg Pro
 1 5

 <210> 46
 35 <211> 8
 <212> PRT
 <213> Methanoregula boonci

 <400> 46
 Gln Met Ala Gly Arg Ala Gly Arg
 40 1 5

 <210> 47
 <211> 9
 <212> PRT
 <213> Methanoregula boonci

 <400> 47
 45 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5

 <210> 48

<211> 8
 <212> PRT
 <213> *Ferroplasma acidarmanus*

 <400> 48
 Gln Met Ile Gly Arg Ala Gly Arg
 5 1 5

 <210> 49
 <211> 9
 <212> PRT
 <213> *Ferroplasma acidarmanus*

 10 <400> 49
 Gln Met Ile Gly Arg Ala Gly Arg Pro
 1 5

 <210> 50
 <211> 8
 <212> PRT
 15 <213> *Methanocaldococcus fervens*

 <400> 50
 Gln Cys Ile Gly Arg Ala Gly Arg
 1 5

 <210> 51
 <211> 9
 20 <212> PRT
 <213> *Methanocaldococcus fervens*

 <400> 51
 Gln Cys Ile Gly Arg Ala Gly Arg Pro
 1 5

 <210> 52
 <211> 8
 <212> PRT
 25 <213> *Methanocaldococcus jannaschii*

 <400> 52
 Gln Cys Ile Gly Arg Ala Gly Arg
 1 5

 30 <210> 53
 <211> 9
 <212> PRT
 <213> *Methanocaldococcus jannaschii*

 <400> 53
 Gln Cys Ile Gly Arg Ala Gly Arg Pro
 35 1 5

 <210> 54
 <211> 8
 <212> PRT
 <213> *Methanocaldococcus infernus*

 40 <400> 54
 Gln Cys Ile Gly Arg Ala Gly Arg
 1 5

 <210> 55
 <211> 9
 <212> PRT
 45 <213> *Methanocaldococcus infernus*

 <400> 55
 Gln Cys Ile Gly Arg Ala Gly Arg Pro
 1 5

<210> 56
 <211> 8
 <212> PRT
 <213> Methanospirillum hungatei
 5 <400> 56
 Gln Met Ala Gly Arg Ala Gly Arg
 1 5
 <210> 57
 <211> 9
 <212> PRT
 10 <213> Methanospirillum hungatei
 <400> 57
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5
 <210> 58
 <211> 8
 15 <212> PRT
 <213> Archaeoglobus fulgidus
 <400> 58
 Gln Met Ala Gly Arg Ala Gly Arg
 1 5
 <210> 59
 <211> 9
 20 <212> PRT
 <213> Archaeoglobus fulgidus
 <400> 59
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5
 25 <210> 60
 <211> 8
 <212> PRT
 <213> Haloterrigena turkmenica
 <400> 60
 Gln Met Ala Gly Arg Ala Gly Arg
 30 1 5
 <210> 61
 <211> 9
 <212> PRT
 <213> Haloterrigena turkmenica
 35 <400> 61
 Gln Met Ala Gly Arg Ala Gly Arg Pro
 1 5
 <210> 62
 <211> 8
 <212> PRT
 40 <213> Haladaptatus paucihalophilus
 <400> 62
 Gln Met Phe Gly Arg Ala Gly Arg
 1 5
 <210> 63
 <211> 9
 45 <212> PRT
 <213> Haladaptatus paucihalophilus
 <400> 63

Gln Met Phe Gly Arg Ala Gly Arg Pro
 1 5
 <210> 64
 <211> 8
 <212> PRT
 5 <213> Halobacterium sp.
 <400> 64
 Gln Met Phe Gly Arg Ala Gly Arg
 1 5
 <210> 65
 <211> 9
 10 <212> PRT
 <213> Halobacterium sp.
 <400> 65
 Gln Met Phe Gly Arg Ala Gly Arg Pro
 1 5
 <210> 66
 <211> 8
 15 <212> PRT
 <213> Secuencia Artificial
 <220>
 <221> fuente
 20 <223> /nota="Descripción de Secuencia Artificial: Péptido sintético"
 <220>
 <221> VARIANT
 <222> (2)..(2)
 <223> /replace="Met" or "Leu"
 25 <220>
 <221> MOD_RES
 <222> (3)..(3)
 <223> Cualquier aminoácido
 <220>
 30 <221> fuente
 <223> /nota="Véase la memoria descriptiva según se presentó para una descripción detallada de sustituciones y realizaciones preferidas"
 <220>
 <221> característica_misc
 35 <222> (1)..(8)
 <223> /nota="Los residuos variantes dados en la secuencia no tienen preferencia con respecto a los de las anotaciones para posiciones variantes"
 <400> 66
 Gln Cys Xaa Gly Arg Ala Gly Arg
 1 5
 40 <210> 67
 <211> 87
 <212> ADN
 <213> Secuencia Artificial
 <220>
 45 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"
 <220>
 <221> base_modificada
 <222> (10)..(10)
 50 <223> nucleótido abásico

<400> 67
 aaaccttccn cccgtaccgt gccgtaccgt tccgttccgt accgtatttt tttttctcac 60
 tatcgcattc tcatgcaggt cgtagcc 87
 <210> 68
 <211> 45
 5 <212> ADN
 <213> Secuencia Artificial
 <220>
 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"
 10 <400> 68
 cctgcagtag aatgcgtag tgagatttt tttttttt tttt 45
 <210> 69
 <211> 102
 <212> ADN
 15 <213> Secuencia Artificial
 <220>
 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: Polinucleótido sintético"
 <220>
 20 <221> base_modificada
 <222> (19)..(20)
 <223> nucleótido abásico
 <400> 69
 25 ctcacctatc cttccactnn ccccttttgg gtttaaattt tttcagatct cactatcttt 60
 ttaaagtttt ctcactatcg cattctcatg caggtcgtag cc 102
 <210> 70
 <211> 91
 <212> ADN
 30 <213> Secuencia Artificial
 <220>
 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"
 <220>
 35 <221> base_modificada
 <222> (19)..(20)
 <223> nucleótido abásico
 <400> 70
 catcatcatc atcatcatnn cccctaaac aagaatacca cgactagcat ttttcagatc 60
 tcactatcgc attctcatgc aggtcgtagc c 91
 40 <210> 71
 <211> 91
 <212> ADN
 <213> Secuencia Artificial
 <220>
 45 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"
 <220>
 <221> base_modificada
 <222> (19)..(20)
 50 <223> nucleótido abásico

<400> 71
catcatcatc atcatcatnn cccctaacc aagaatacca cgactagcat ttttcagatc 60
tcactatcgc attctcatgc aggtcgtagc c 91

<210> 72
<211> 87
5 <212> ADN
<213> Secuencia Artificial

<220>
<221> fuente
<223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"

10 <220>
<221> base_modificada
<222> (10)..(10)
<223> nucleótido abásico

<400> 72
aaaccttccn cccgtaccgt gccgtaccgt tccgttccgt accgtatttt tttttctcac 60
15 tatcgcattc tcatgcaggt cgtagcc 87

<210> 73
<211> 60
<212> ADN
<213> Secuencia Artificial

20 <220>
<221> fuente
<223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"

<400> 73
aaaaaaaata cgtacggaa cggaacggta cggcacggta cgggttttt tttttttt 60

25 <210> 74
<211> 14
<212> ADN
<213> Secuencia Artificial

30 <220>
<221> fuente
<223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"

<400> 74
atcatcatca tcat 14

35 <210> 75
<211> 9
<212> PRT
<213> Desconocido

<220>
<221> fuente
40 <223> /nota="Descripción de Desconocido: "secuencia de resto de familia 'SF2'"

<400> 75
Gly Phe Cys Cys Pro Gln Ser Ile Gln
1 5

<210> 76
<211> 8
45 <212> PRT
<213> Desconocido

<220>
<221> fuente
<223> /nota="Descripción de Desconocido: "secuencia de resto de familia 'SF2'"

	<220> <221> MOD_RES <222> (2)..(3) <223> Cualquier aminoácido
5	<220> <221> MOD_RES <222> (5)..(5) <223> Cualquier aminoácido
10	<400> 76 Ala Xaa Xaa Gly Xaa Gly Lys Thr 1 5
	<210> 77 <211> 6 <212> PRT <213> Desconocido
15	<220> <221> fuente <223> /nota="Descripción de Desconocido: "secuencia de resto de familia 'SF2"
	<400> 77 Pro Thr Arg Glu Leu Ala 1 5
20	<210> 78 <211> 4 <212> PRT <213> Desconocido
25	<220> <221> fuente <223> /nota="Descripción de Desconocido: "secuencia de resto de familia 'SF2"
	<400> 78 Thr Pro Gly Arg 1
30	<210> 79 <211> 6 <212> PRT <213> Desconocido
35	<220> <221> fuente <223> /nota="Descripción de Desconocido: "secuencia de resto de familia 'SF2"
	<400> 79 Val Leu Asp Glu Ala Asp 1 5
40	<210> 80 <211> 4 <212> PRT <213> Desconocido
45	<220> <221> fuente <223> /nota="Descripción de Desconocido: "secuencia de resto de familia 'SF2"
	<400> 80 Phe Val Asn Thr 1
	<210> 81 <211> 8 <212> PRT

<213> Desconocido

<220>

<221> fuente

<223> /nota="Descripción de Desconocido: "secuencia de resto de familia 'SF2'

5 <220>

<221> MOD_RES

<222> (6)..(7)

<223> Cualquier aminoácido

<400> 81

His Arg Ile Gly Arg Xaa Xaa Arg

10 1 5

<210> 82

<211> 18

<212> ADN

<213> Secuencia Artificial

15 <220>

<221> fuente

<223> /nota="Descripción de Secuencia Artificial: Cebador sintético"

<400> 82

tctcgctgct catcaact 18

20 <210> 83

<211> 18

<212> ADN

<213> Secuencia Artificial

<220>

25 <221> fuente

<223> /nota="Descripción de Secuencia Artificial: Cebador sintético"

<400> 83

gagtcgaggt catatcgt 18

<210> 84

30 <211> 18

<212> ADN

<213> Secuencia Artificial

<220>

<221> fuente

35 <223> /nota="Descripción de Secuencia Artificial: Cebador sintético"

<400> 84

gtctgcctat agtgagtc 18

<210> 85

<211> 65

40 <212> ADN

<213> Secuencia Artificial

<220>

<221> fuente

<223> /nota="Descripción de Secuencia Artificial: Cebador sintético"

45 <220>

<221> base_modificada

<222> (22)..(23)

<223> nucleótido abásico

<400> 85

tttttttttt tttttccttc cnnnttttctt cttcttcaag aagaagatct cgtcgctcat 60

50 caact 65

<210> 86
 <211> 65
 <212> ADN
 <213> Secuencia Artificial

5 <220>
 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: Cebador sintético"

<220>
 <221> base_modificada
 10 <222> (22)..(23)
 <223> nucleótido abásico

<400> 86
 tttttttttt ttttttcttc cnntttttta attaatattg ttgttgtgag tgcaggtcat 60
 atcgt 65

15 <210> 87
 <211> 41
 <212> ADN
 <213> Secuencia Artificial

<220>
 <221> fuente
 20 <223> /nota="Descripción de Secuencia Artificial: Cebador sintético"

<220>
 <221> fuente
 <223> /nota="Descripción de Molécula de ADN/ARN combinada: Cebador sintético"

25 <220>
 <221> base_modificada
 <222> (9)..(9)
 <223> deoxiuracilo

30 <220>
 <221> base_modificada
 <222> (19)..(19)
 <223> deoxiuracilo

35 <220>
 <221> base_modificada
 <222> (27)..(27)
 <223> deoxiuracilo

40 <220>
 <221> base_modificada
 <222> (36)..(36)
 <223> deoxiuracilo

<400> 87
 atacggcgnc caccgaccnc agcgtcngcc tatagngagt c 41

45 <210> 88
 <211> 29
 <212> ADN
 <213> Secuencia Artificial

<220>
 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"

50 <400> 88
 accgacactg cgtctgccta tagtgagtc 29

<210> 89
 <211> 20

<212> ADN
 <213> Secuencia Artificial

 <220>
 <221> fuente
 5 <223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"

 <400> 89
 ttaattaatt ttgtgtgt 20

 <210> 90
 <211> 20
 10 <212> ADN
 <213> Secuencia Artificial

 <220>
 <221> fuente
 <223> /nota="Descripción de Secuencia Artificial: Oligonucleótido sintético"

 15 <400> 90
 cttcttcttc aagaagaaga 20

REIVINDICACIONES

1. Un método de caracterización de un polinucleótido diana, comprendiendo el método:

(a) aplicar una diferencia de potencial a través de un poro en contacto con una helicasa Hel308 y el polinucleótido diana;

5 (b) medir al menos dos señales de diferentes magnitudes por nucleótido del polinucleótido diana que se mueve a través del poro durante un ciclo de translocación completo de la helicasa Hel308; y

(c) caracterizar el polinucleótido diana utilizando al menos dos señales de diferentes magnitudes por nucleótido medido en (b).

2. El método de la reivindicación 1, que comprende además:

10 (d) después de (c), variar al menos un parámetro a fin de variar la temporización de una o más etapas de translocación fraccional por la helicasa Hel308 del polinucleótido diana a través del poro; y

(e) repetir las etapas (a)-(c) utilizando al menos un parámetro variado.

3. El método de la reivindicación 1, caracterizado porque la caracterización comprende la aplicación de un algoritmo de Viterbi modificado; o

15 caracterizado porque la caracterización comprende niveles de detección e identificación en al menos dos señales y la determinación y resolución de una secuencia del polinucleótido diana basado en los niveles detectados e identificados; opcionalmente en el que la detección e identificación de los niveles en una o más señales incluye la resolución de uno o más de los niveles completos, niveles fraccionales, todos los niveles, y los identificadores de nivel; opcionalmente en donde:

20 la determinación y resolución de la secuencia del polinucleótido diana basado en los niveles detectados e identificados comprende tomar como entrada uno o más de los niveles completos, niveles fraccionales, todos los niveles, y los identificadores de nivel, resolviendo una pluralidad de secuencias basadas en la entrada, y seleccionando y dando resolución a al menos una de las secuencias resueltas basadas en la información básica respecto a la pluralidad de las secuencias resueltas;

25 la determinación y resolución de la secuencia del polinucleótido diana basado en los niveles detectados e identificados comprende tomar como entrada uno o más de los niveles completos, niveles fraccionales, todos los niveles, y los identificadores de nivel, resolviendo una pluralidad de secuencias basadas en la entrada, y seleccionando y concatenando con otras porciones de una pluralidad de las secuencias resueltas basadas en la información básica sobre las porciones de la pluralidad de las secuencias resueltas;

30 la determinación y resolución de la secuencia del polinucleótido diana basado en los niveles detectados e identificados comprende tomar como entrada uno o más de los niveles completos, niveles fraccionales, todos los niveles y los identificadores de nivel, resolviendo una pluralidad de secuencias en función de la entrada, comparando las secuencias resueltas con las secuencias modelo y seleccionando y dando resolución a al menos una de las secuencias resueltas en función de la información básica respecto a la comparación de la secuencia resuelta con la secuencia del modelo; o

35 la determinación y resolución de la secuencia del polinucleótido diana basado en los niveles detectados e identificados comprende tomar como entrada uno o más de los niveles completos, niveles fraccionales, todos los niveles, y los identificadores de nivel, resolviendo una pluralidad de secuencias basadas en la entrada, comparando las secuencias resueltas con las secuencias modelo, y seleccionando y concatenando con otras porciones de una pluralidad de las secuencias resueltas basadas en la información básica respecto a la comparación de porciones de la pluralidad de secuencias resueltas con la secuencia modelo.

40

4. El método de la reivindicación 1, que comprende:

poner en contacto dicha helicasa Hel308 con una concentración de un sustrato de helicasa Hel308 que es diferente de una concentración de referencia de dicho sustrato, produciendo la concentración de sustrato un cambio en la duración de una etapa de translocación fraccional proporcional a una diferencia en la concentración del sustrato en comparación con dicha concentración de referencia.

45

5. El método de la reivindicación 4, caracterizado porque la concentración de sustrato es una concentración sub-saturante del sustrato de helicasa Hel308;

caracterizado porque la concentración de referencia es una concentración saturante del sustrato de helicasa Hel308;

50

caracterizado porque tanto el sustrato como las concentraciones de referencia no son concentraciones saturantes

del sustrato;

caracterizado porque la concentración de sustrato y las concentraciones de referencia son concentraciones sub-saturantes del sustrato de helicasa Hel308; o

en el que el sustrato de helicasa Hel308 es trifosfato de adenosina (ATP).

5 6. El método de la reivindicación 1 o la reivindicación 4 caracterizado porque la caracterización del polinucleótido diana comprende identificar una o más de: una secuencia del polinucleótido diana, una modificación del polinucleótido diana, una longitud del polinucleótido diana, una identidad del polinucleótido diana, una fuente del polinucleótido diana y una estructura secundaria del polinucleótido diana.

10 7. El método de la reivindicación 1 o la reivindicación 4, caracterizado porque la diferencia de potencial comprende una diferencia de potencial eléctrico.

8. El método de la reivindicación 1 o la reivindicación 4, caracterizado porque al menos dos señales comprenden una señal óptica; o

en el que al menos dos señales comprenden una señal eléctrica; opcionalmente en las que:

15 cada una de las dos señales eléctricas al menos es una medida seleccionada a partir de corriente, voltaje, efecto túnel, resistencia, potencial, voltaje, conductancia y medición eléctrica transversal; opcionalmente en donde

cada una de las dos señales eléctricas, por lo menos, comprende una corriente eléctrica que pasa a través del poro,

o

20 caracterizado porque una unidad de uno o más nucleótidos del polinucleótido diana que pasa a través del poro durante un ciclo de translocación completo de la helicasa Hel308 se caracteriza por una precisión superior al 50% utilizando al menos dos señales medidas en (b), en comparación con la caracterización de la unidad de uno o más nucleótidos del polinucleótido diana utilizando una sola señal medida durante el ciclo de translocación completo.

9. El método de la reivindicación 1 o la reivindicación 4, caracterizado porque el ciclo de translocación completo de la helicasa Hel308 comprende un primer etapa de translocación fraccional; y

25 una segunda etapa de translocación fraccional, y

caracterizado porque una translocación del polinucleótido diana se encuentra en una dirección opuesta a la fuerza aplicada por la diferencia potencial en el polinucleótido que se transloca a través del poro; o

en el que una translocación del polinucleótido diana está en una dirección con la fuerza aplicada proporcionada por la diferencia de potencial en el polinucleótido que se transloca a través del poro.

30 10. El método de la reivindicación 1 o la reivindicación 4, caracterizado porque uno o más residuos de nucleótidos en el polinucleótido diana se caracterizan por una mayor precisión a una concentración inferior del sustrato en comparación con la concentración de referencia.

11. El método de la reivindicación 1 o la reivindicación 4, caracterizado porque el poro es un poro biológico; opcionalmente en donde:

35 el poro biológico es un poro de polinucleótido; o

el poro biológico es un poro de polipéptido; opcionalmente en donde:

el poro de polipéptido tiene una zona de constricción de cinco nucleótidos o menos; o el poro de polipéptido comprende una porina A de *Mycobacterium smegmatis* (MspA); opcionalmente en donde

40 el MspA tiene una secuencia de aminoácidos de SEC ID NO: 1 o tiene al menos un 70%, al menos 75%, al menos 80%, al menos 85%, al menos 90%, al menos 95%, o al menos 99% de homología con SEC ID NO: 1.

12. El método de la reivindicación 1 o la reivindicación 4, caracterizado porque el poro es un poro en estado sólido; o

en el que el poro es un poro híbrido en estado sólido y biológico; opcionalmente, en el que el poro híbrido biológico y sólido es un poro híbrido en estado sólido polipéptido; o

el poro híbrido en estado sólido y biológico es un poro híbrido en estado sólido de polinucleótido.

45 13. El método de la reivindicación 1 o la reivindicación 4, caracterizado porque la helicasa Hel308 es una helicasa mostrada en las tablas 1 y 2 o una variante de la misma.

14. El método de la reivindicación 1 o la reivindicación 4, caracterizado porque el polinucleótido diana se selecciona del grupo constituido por un único polinucleótido monocatenario, un polinucleótido bicatenario y un polinucleótido parcialmente bicatenario.
- 5 15. El método de la reivindicación 2 comprende además combinar las señales producidas durante (c) y (e) y caracterizar el polinucleótido diana basado en las señales combinadas.
16. El método de la reivindicación 15, caracterizado porque el al menos un parámetro variado se selecciona del grupo constituido por la temperatura, la concentración de sal, la concentración de cofactor, la concentración del producto ATP, el pH y la helicasa Hel308 en particular utilizada.

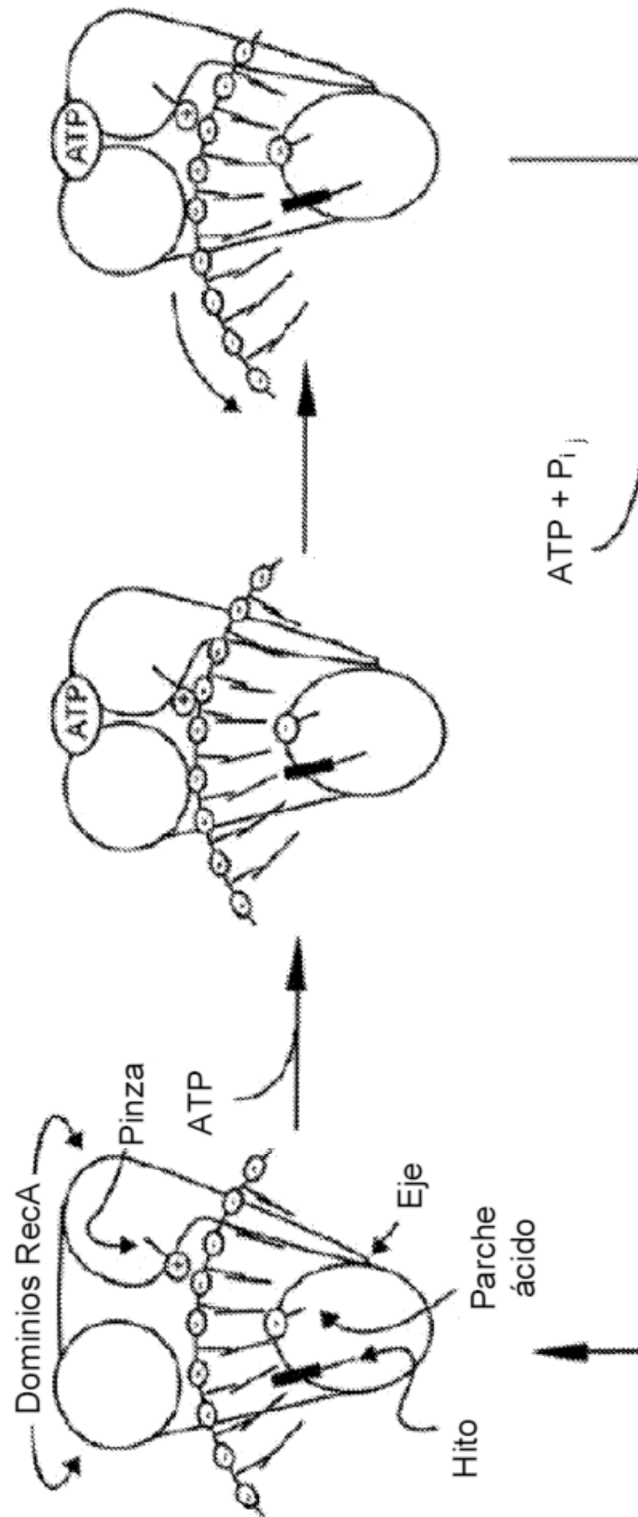


Figura 1A

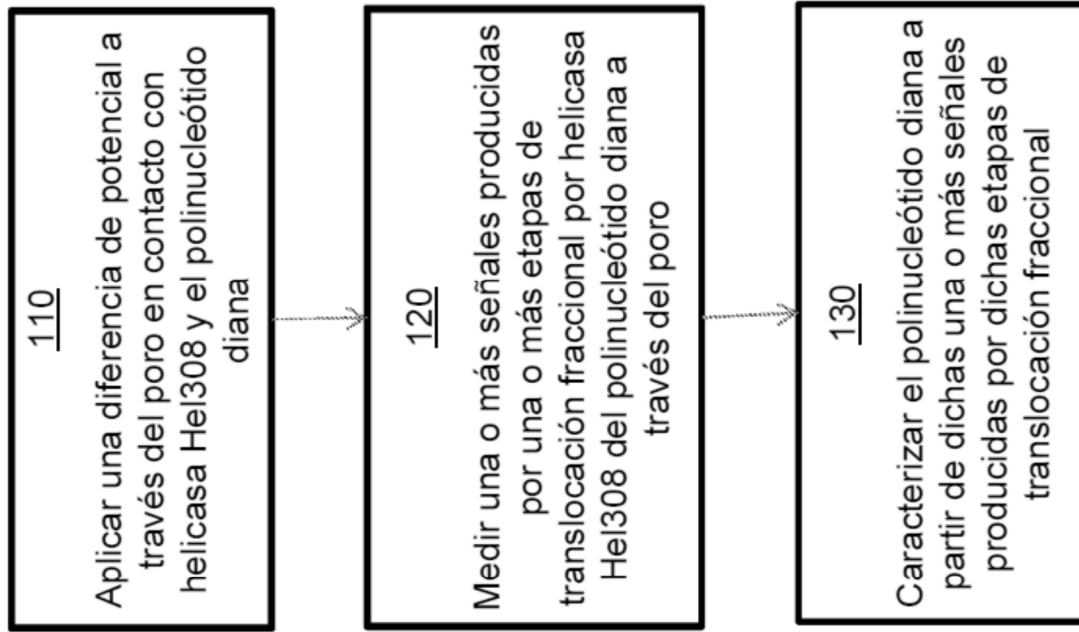


Figura 1C

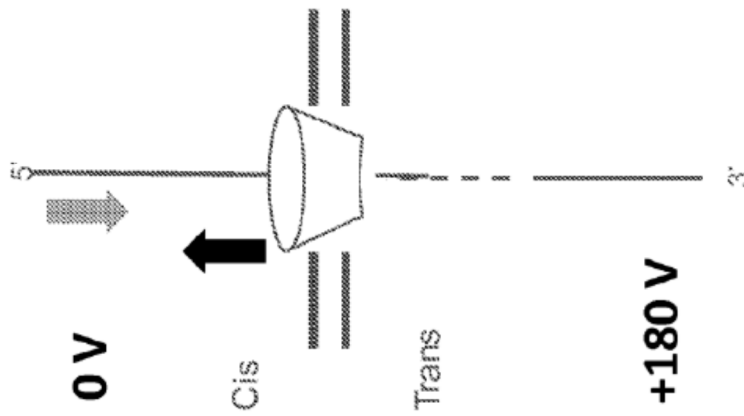


Figura 1B

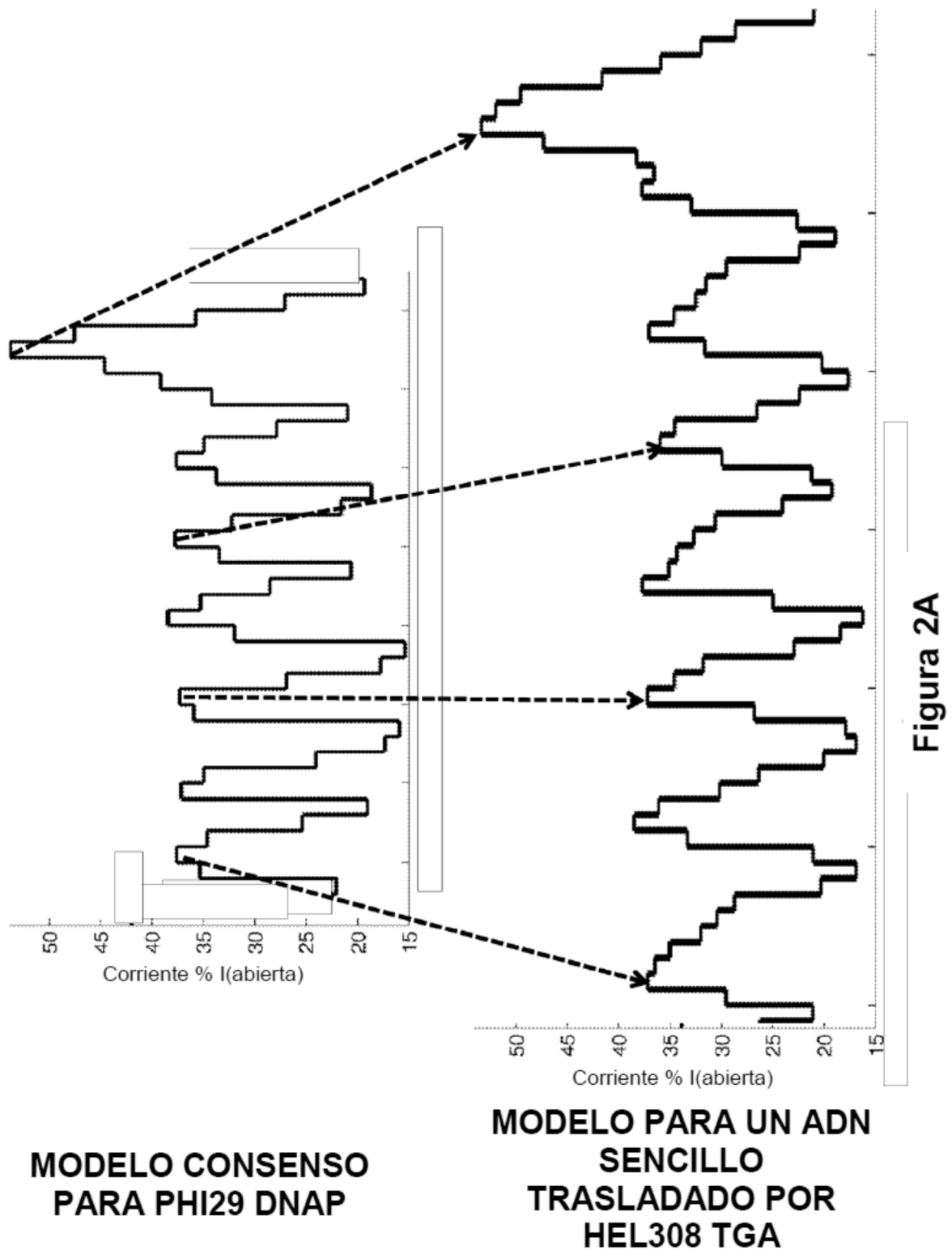


Figura 2A

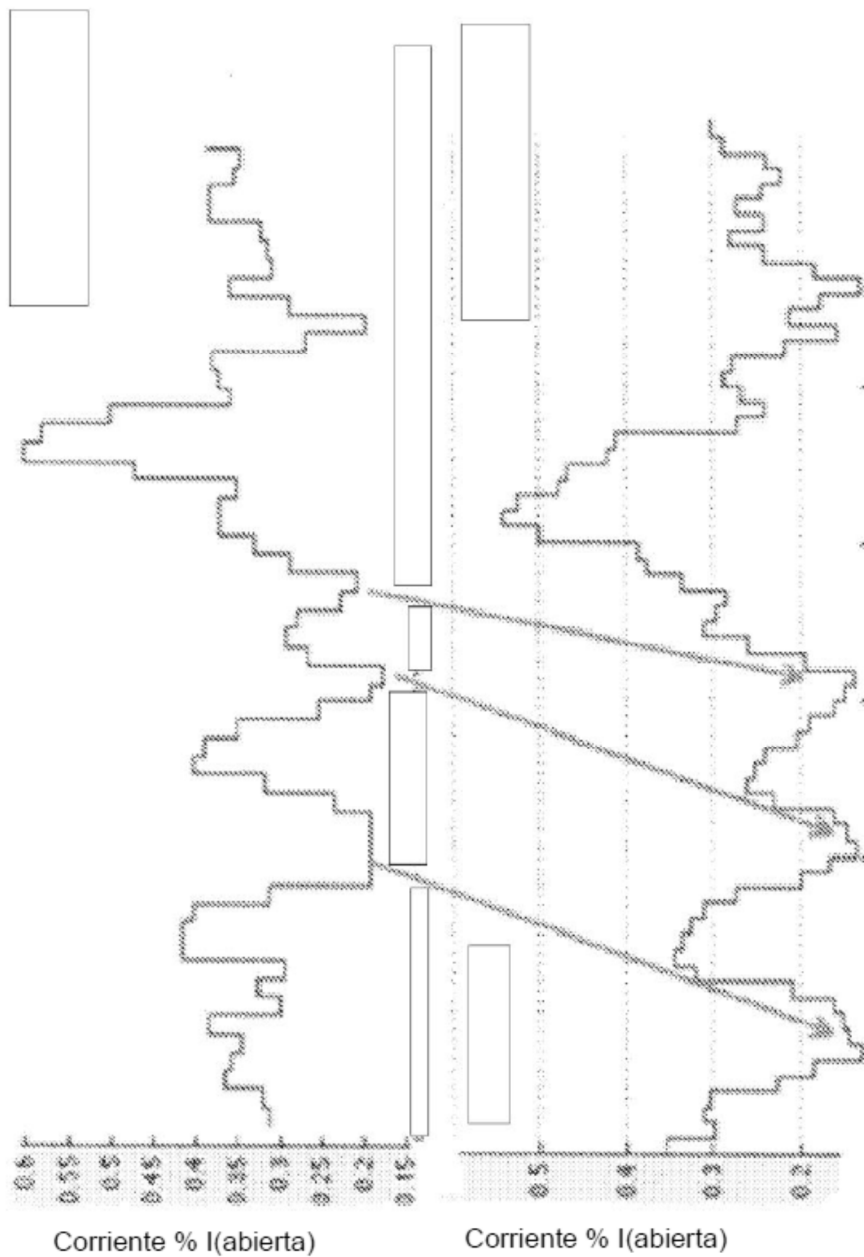


Figura 2B

**MODELO PREDICHO
PARA PHI29 DNAP**

**MODELO PARA UN ADN
SENCILLO
TRASLADADO POR
HEL308 TGA**

3' T A C T A C T A C T A C T A C T A C T

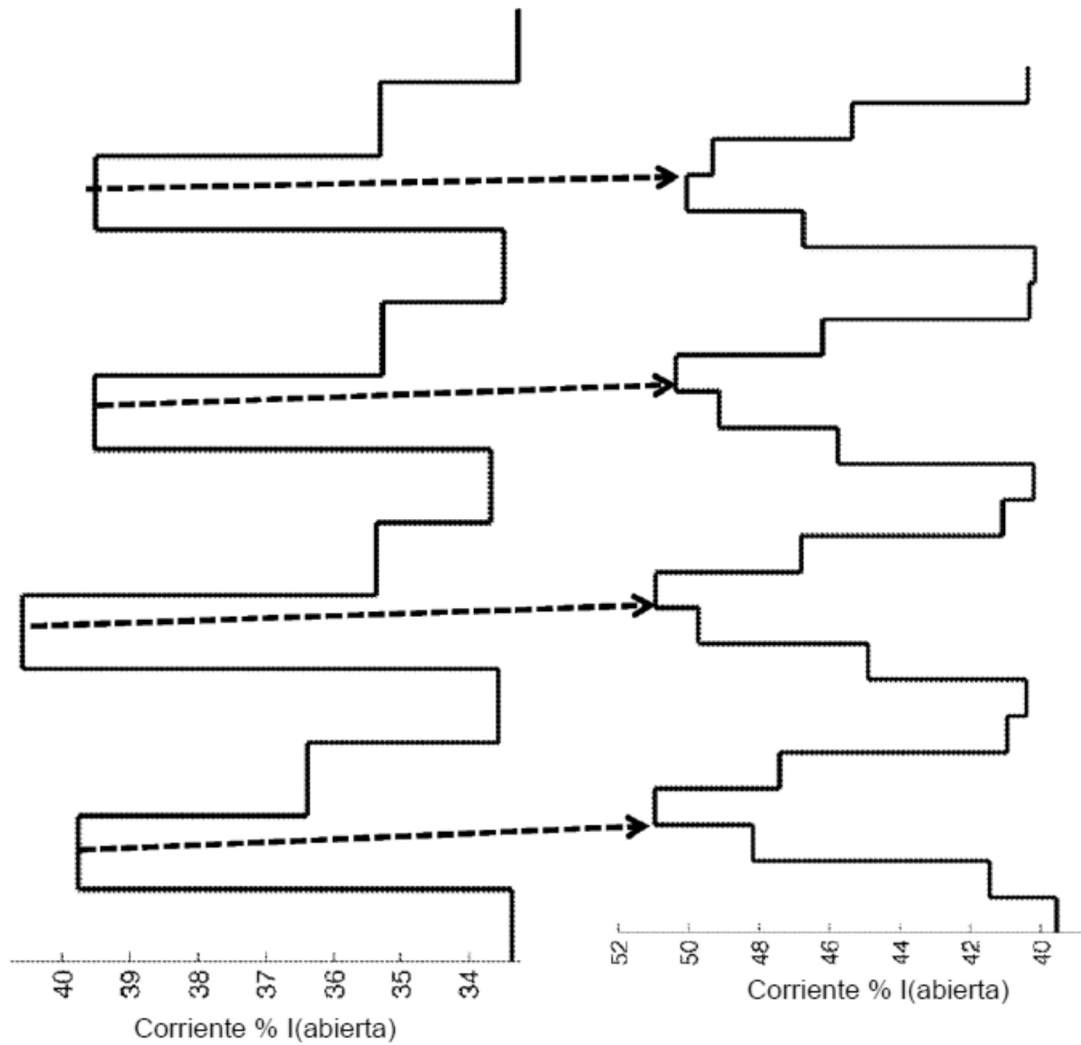


Figura 2C

**MODELO CONSENSO
PARA PHI29 DNAP**

**MODELO CONSENSO
PARA HEL308 TGA**

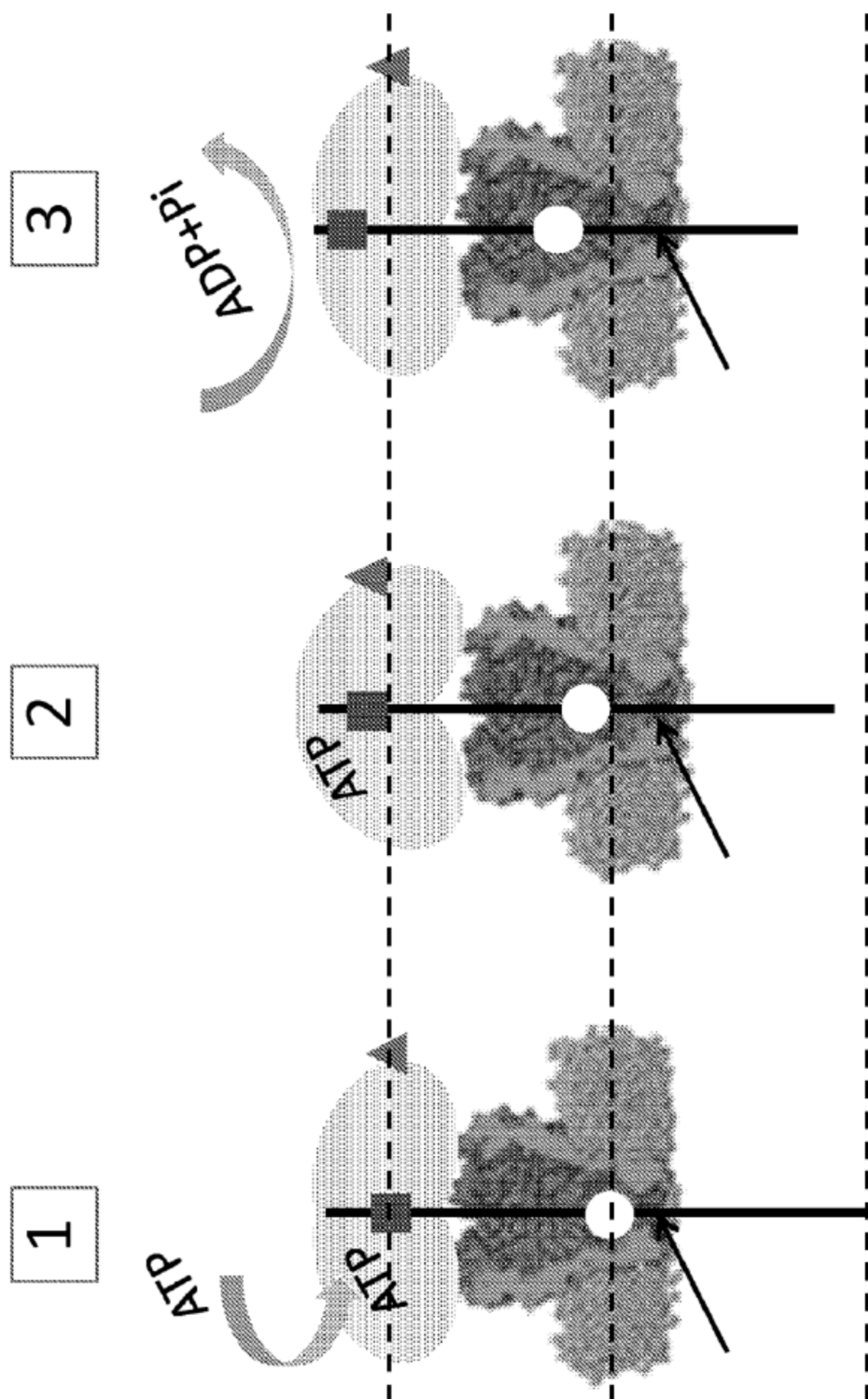


Figura 3

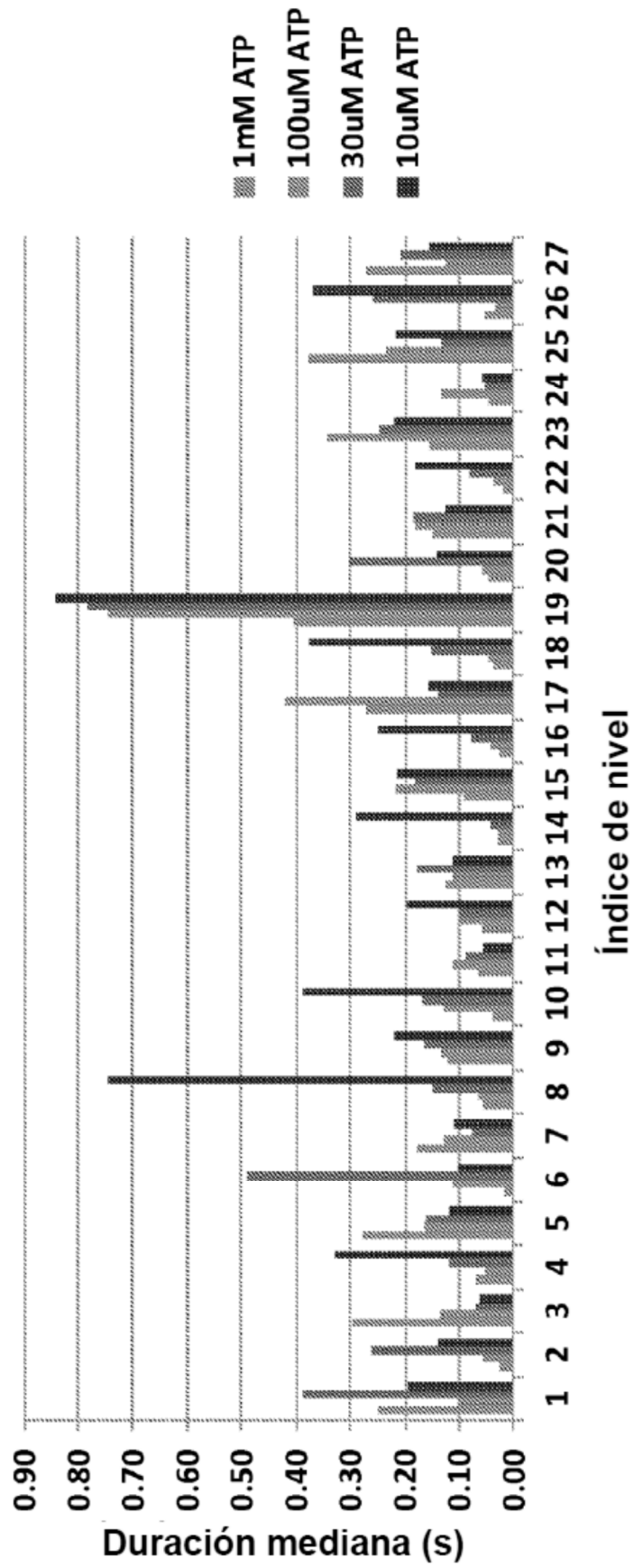


Figura 4A

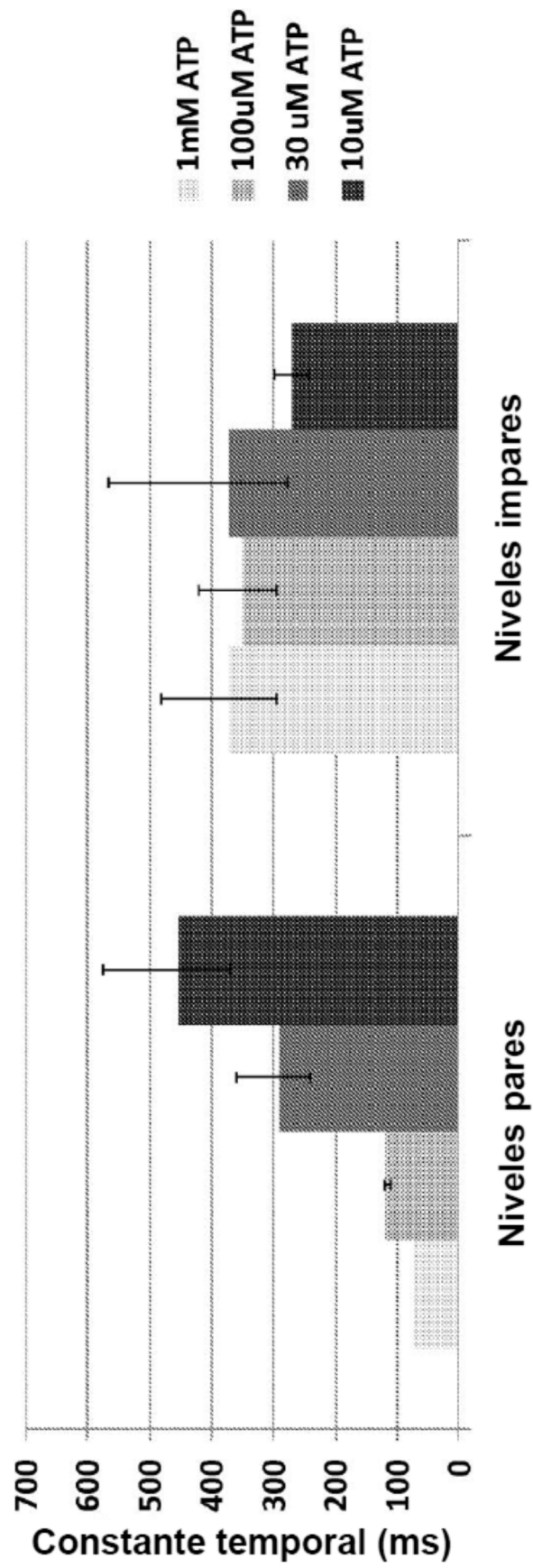


Figura 4B

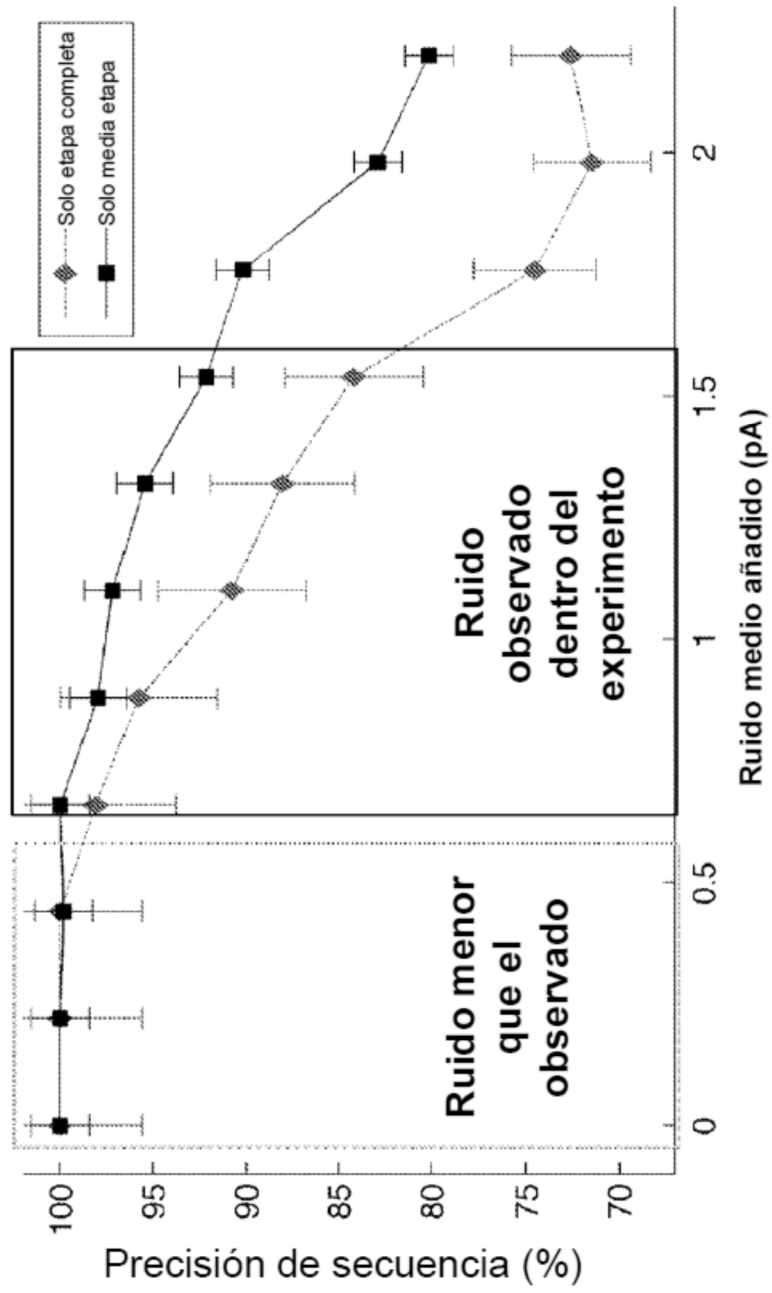


Figura 5

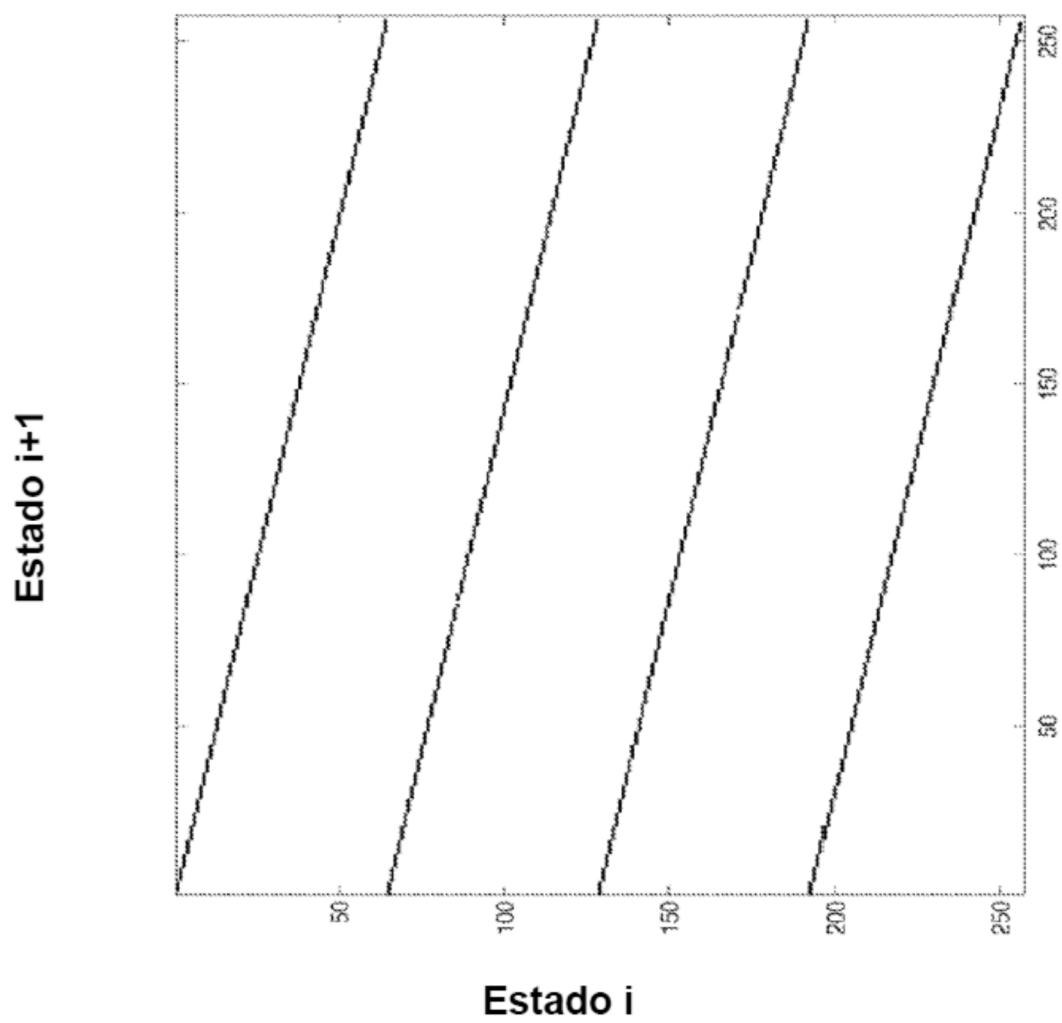


Figura 6A

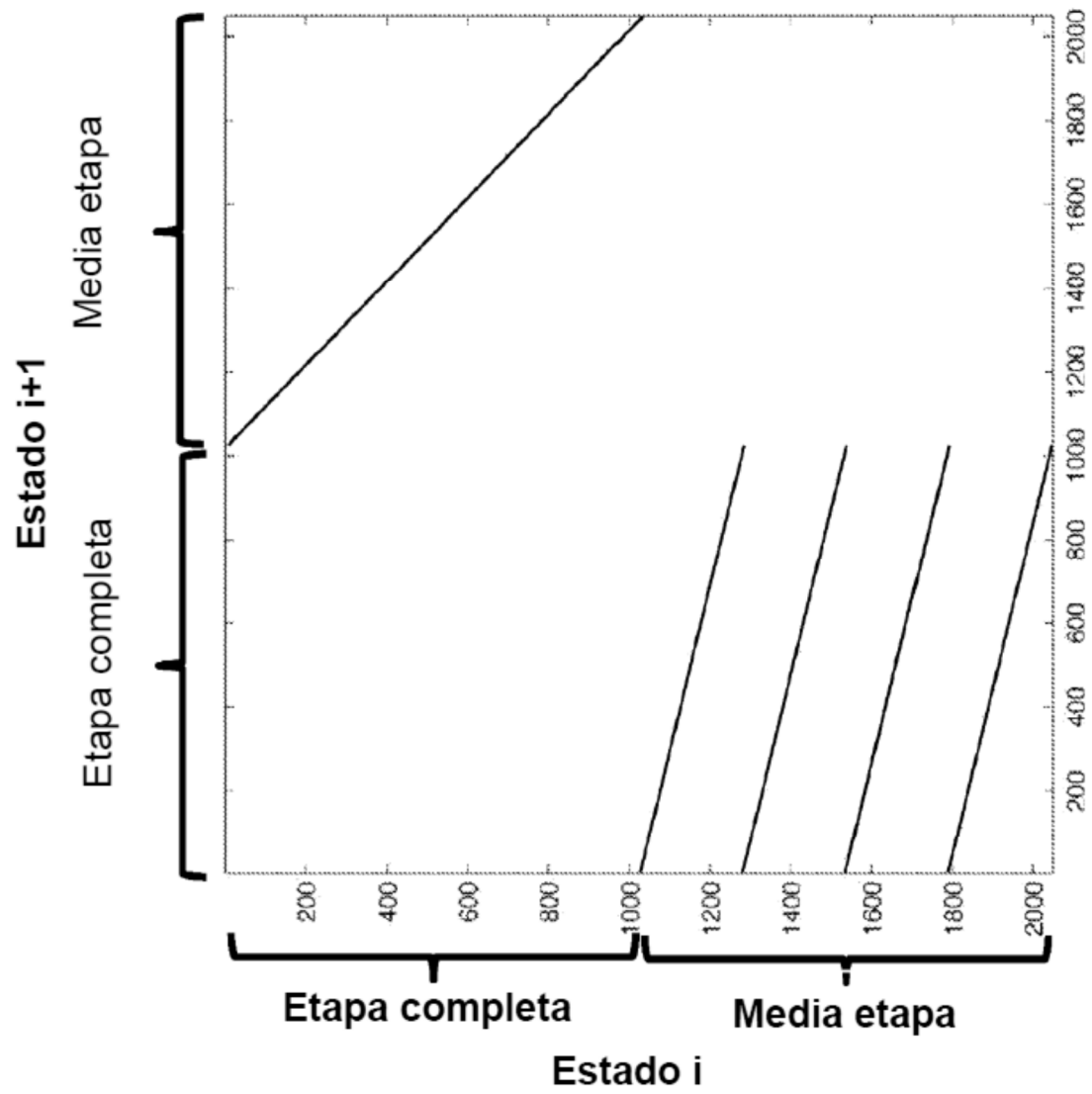
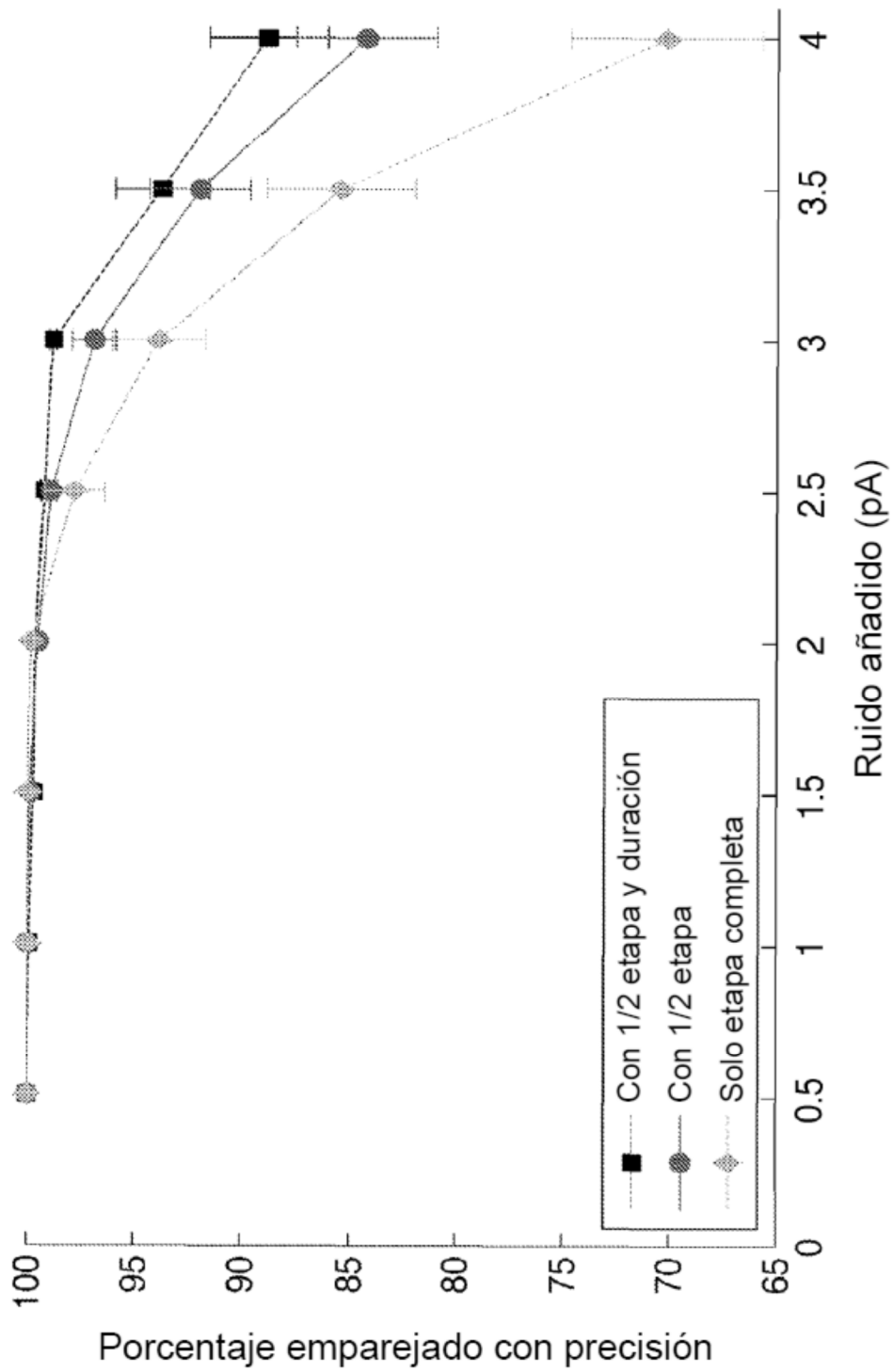


Figura 6B

**Figura 7**

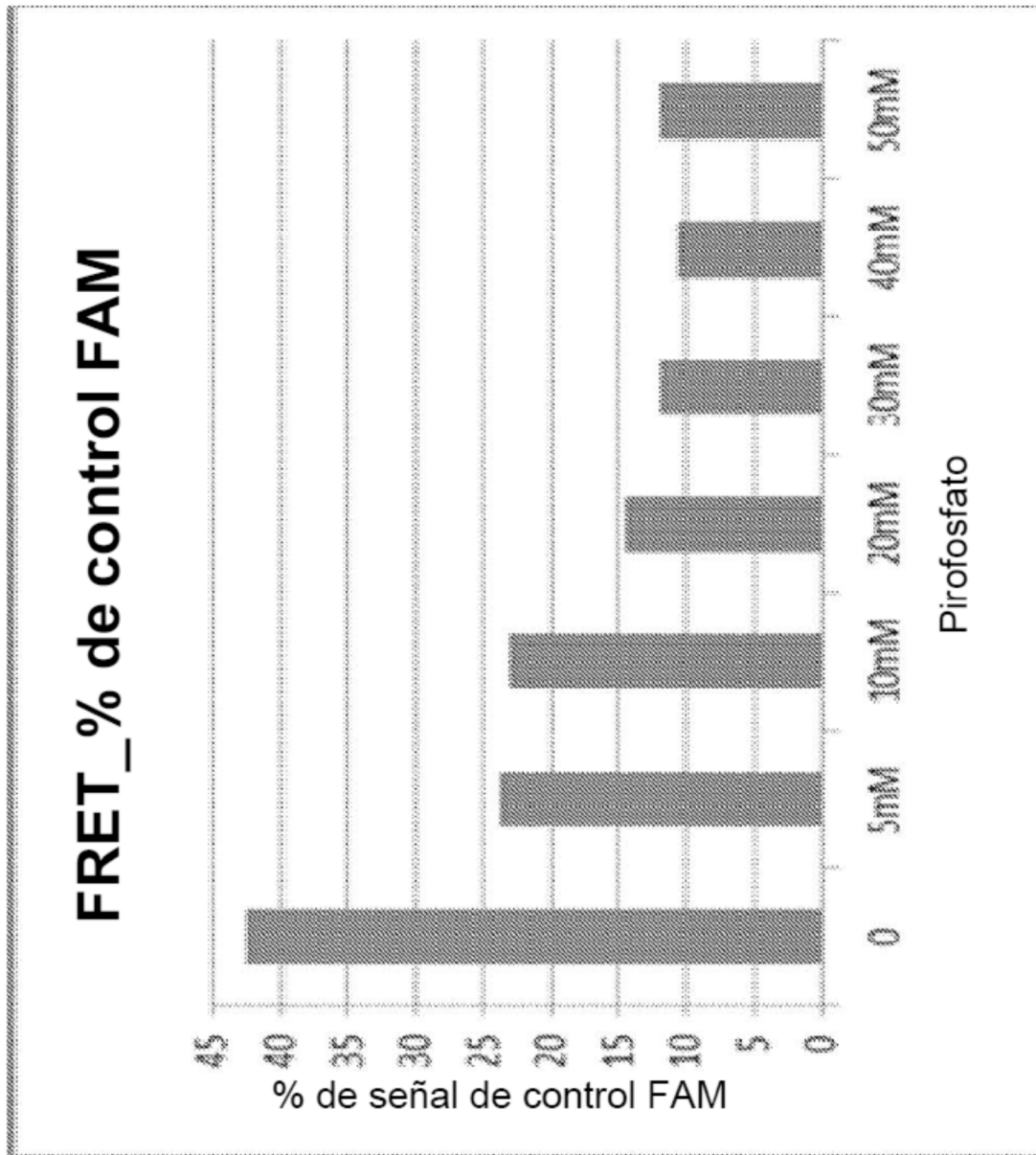


Figura 8

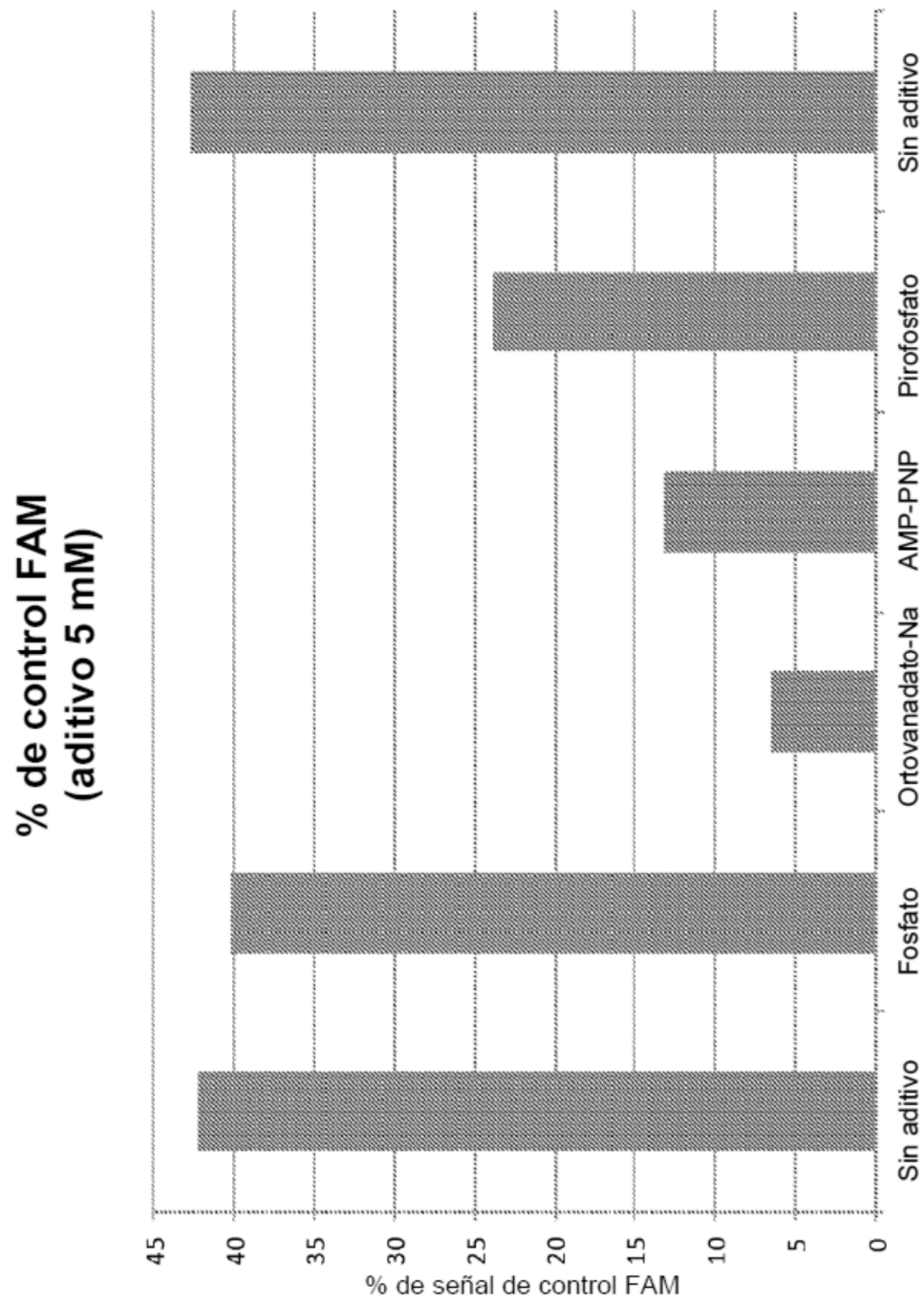


Figura 9

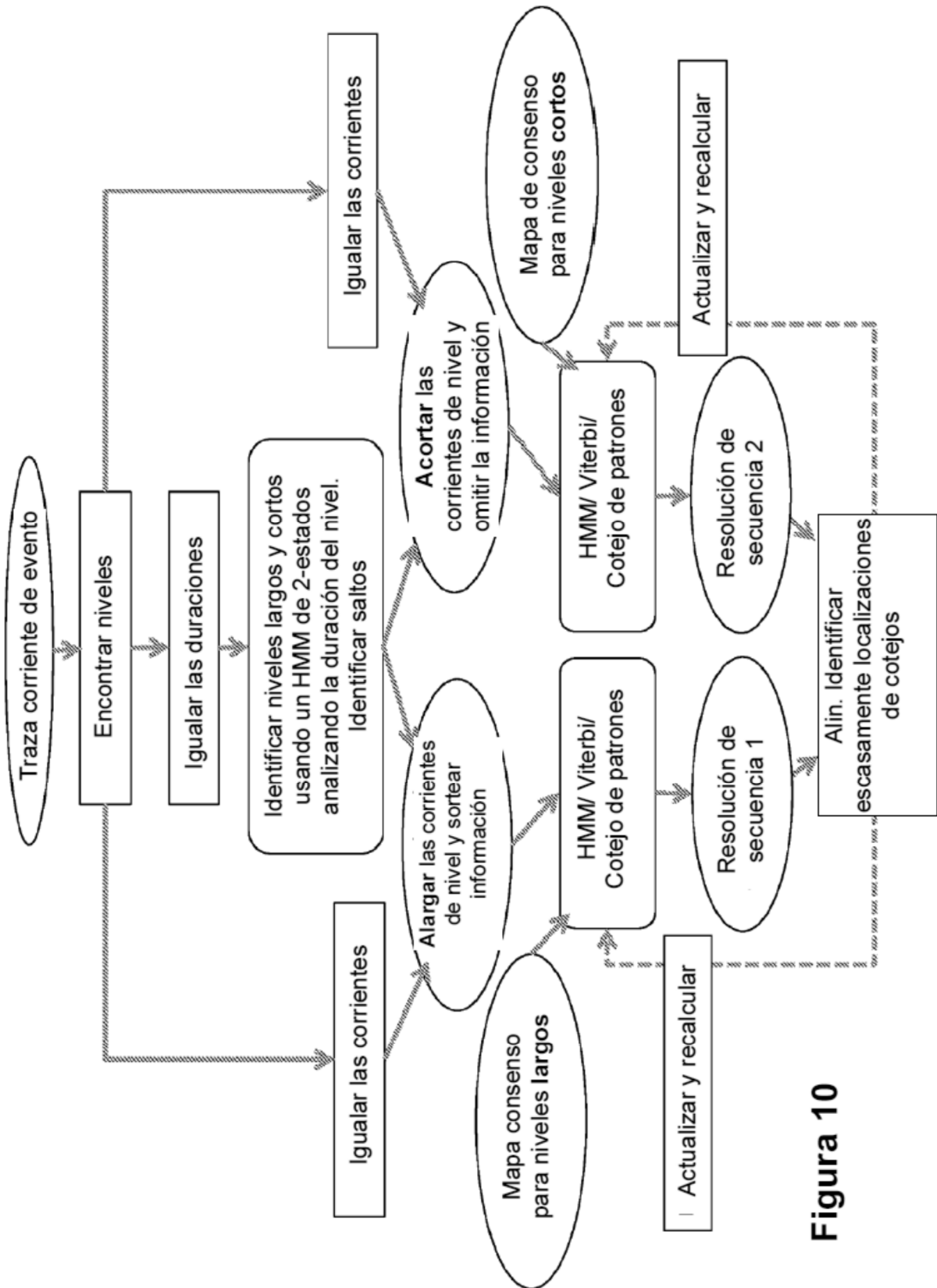


Figura 10

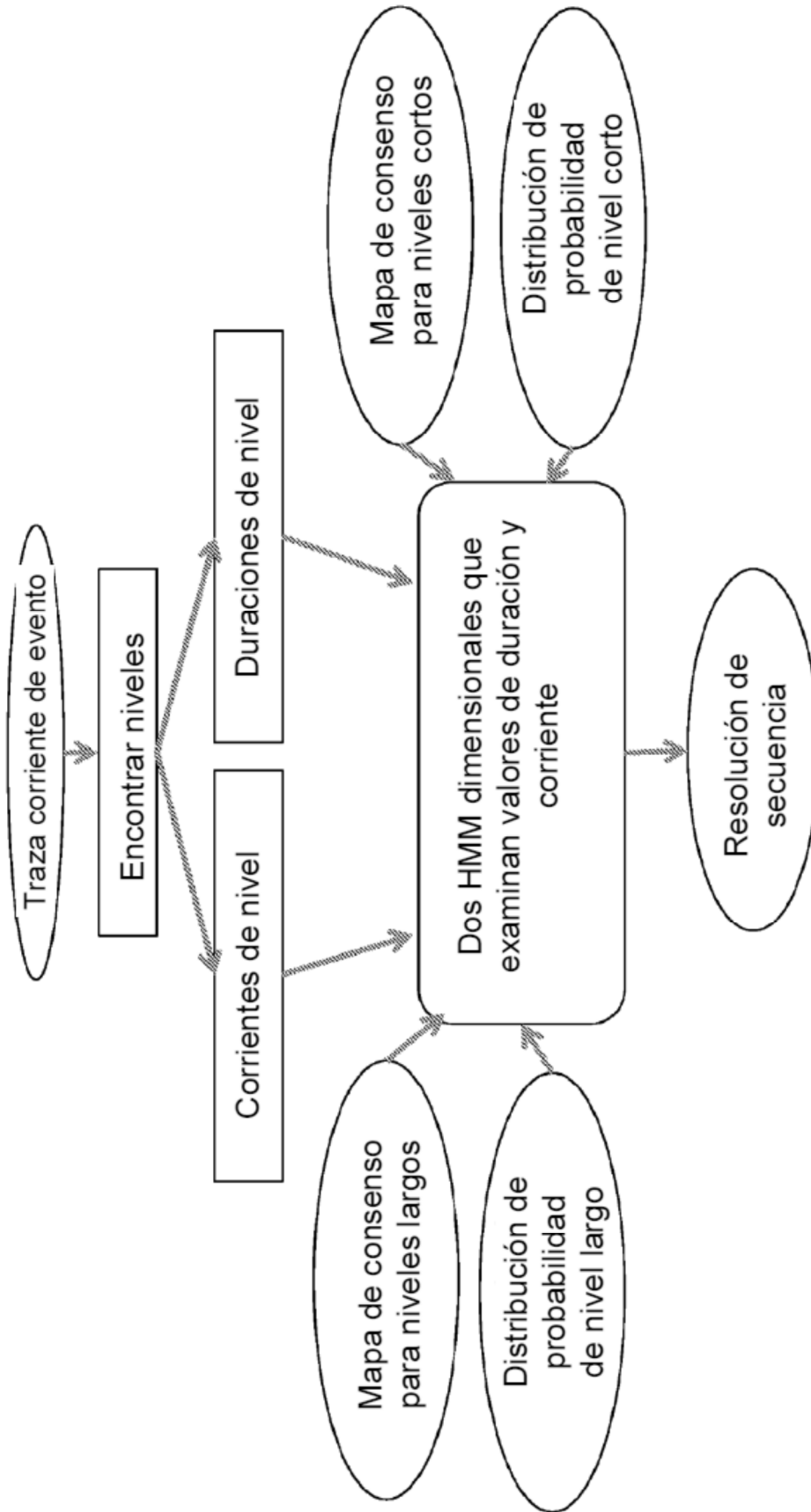


Figura 11

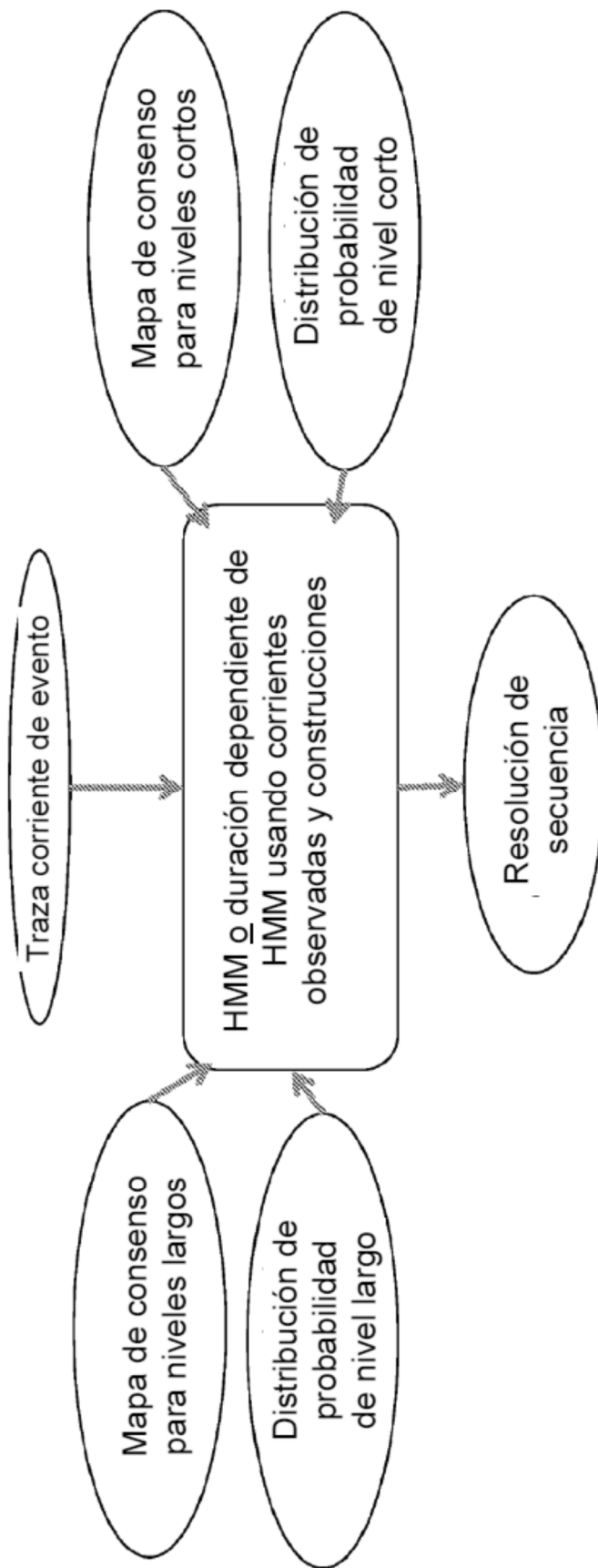
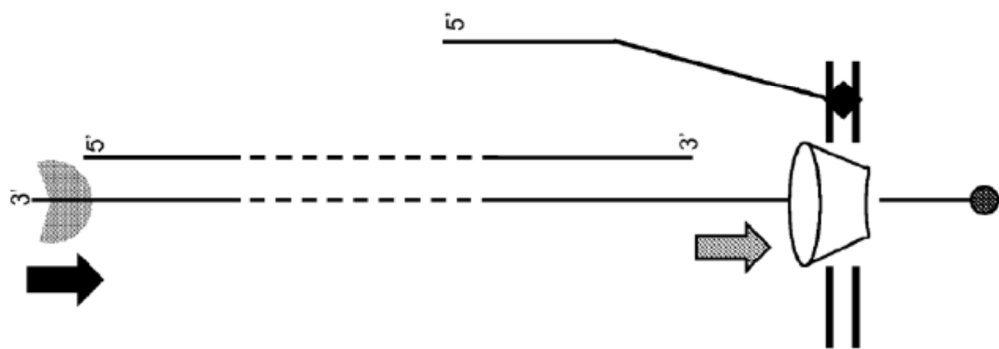
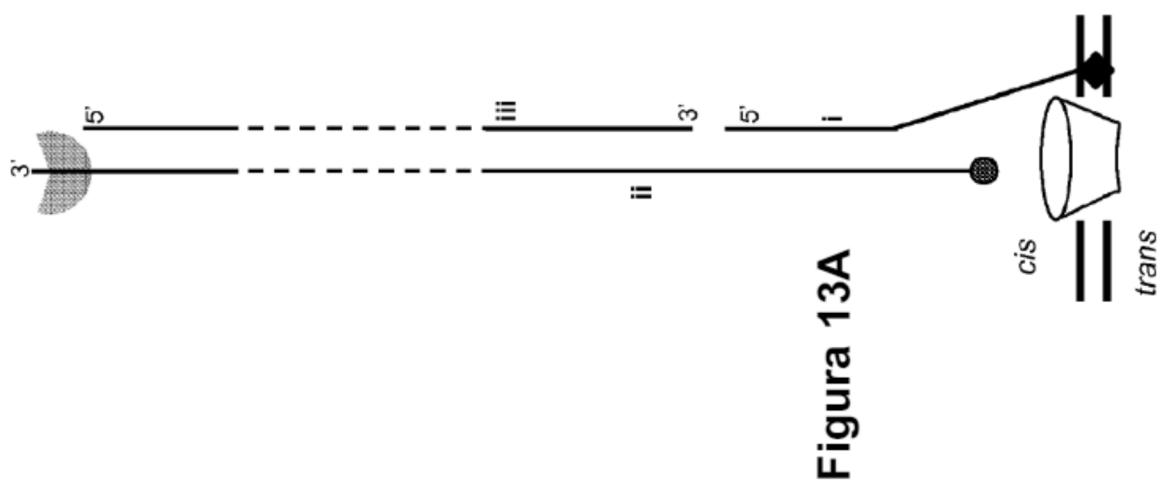


Figura 12



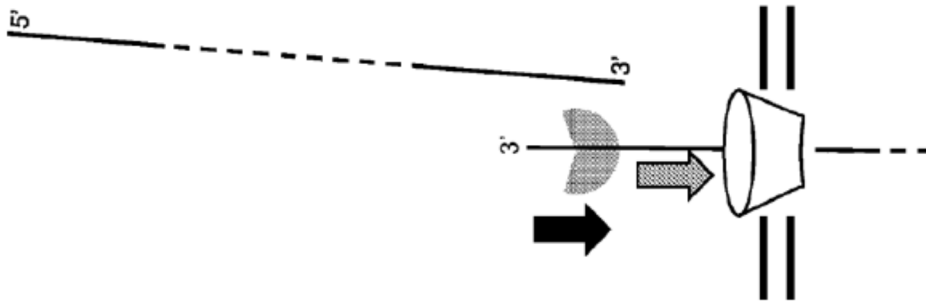


Figura 13C

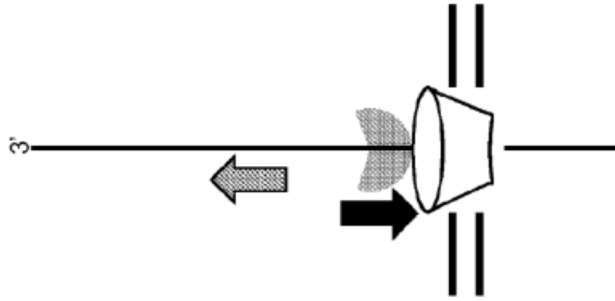


Figura 13D



Figura 13E

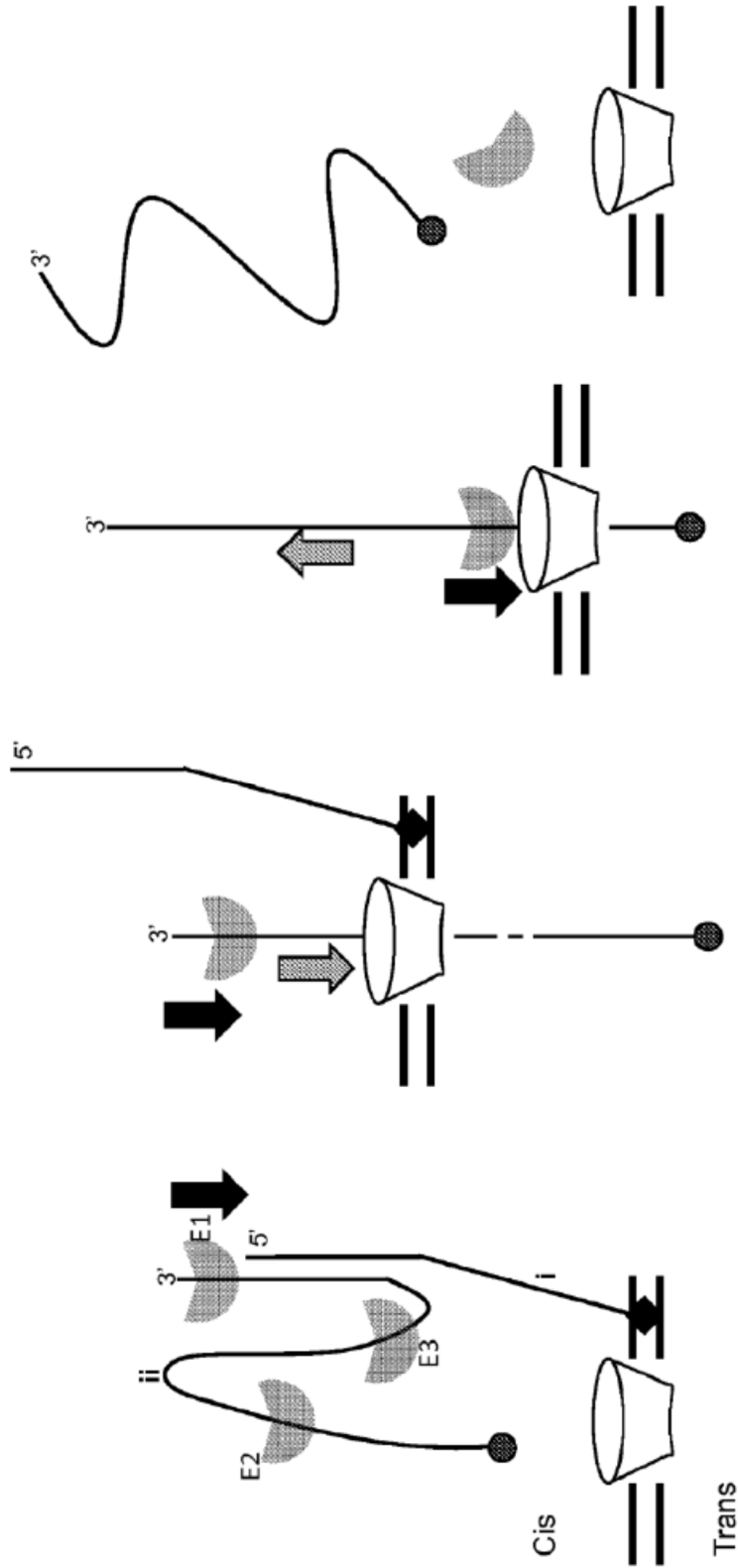


Figura 14D

Figura 14C

Figura 14B

Figura 14A

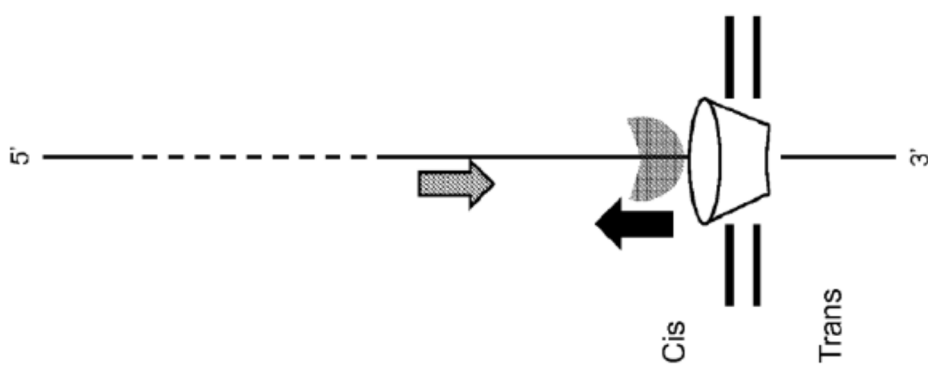


Figura 15A

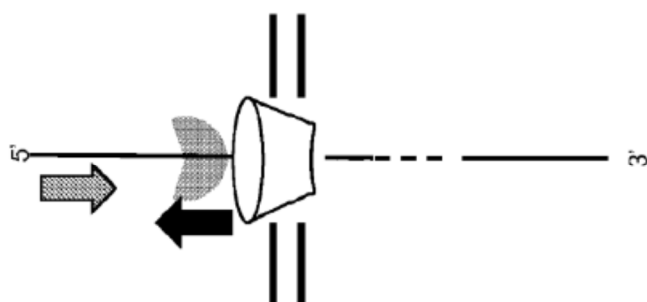


Figura 15B

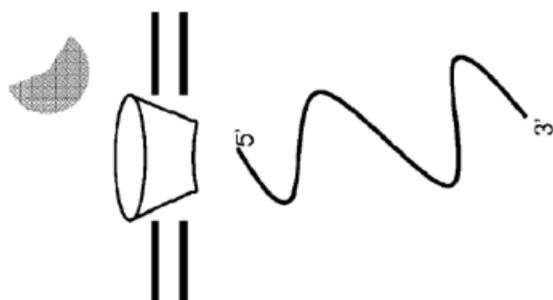


Figura 15C

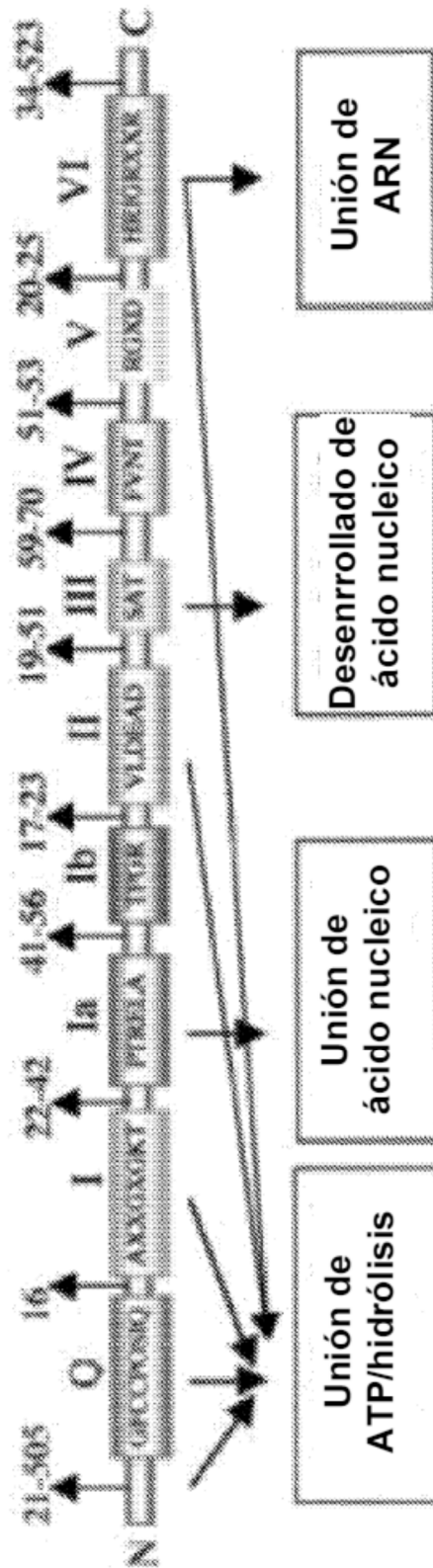


Figura 16. Adaptado de Tuteja et al.

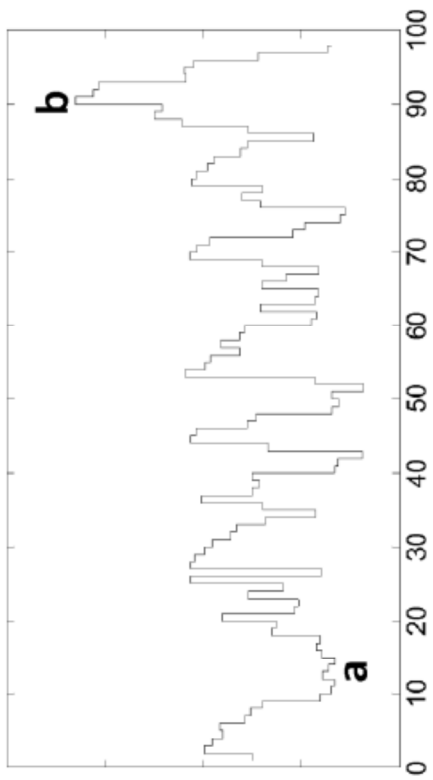


Figura 17B

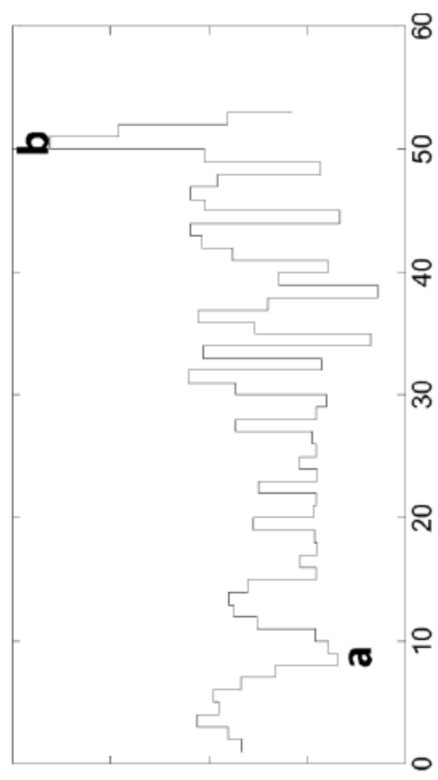


Figura 17D

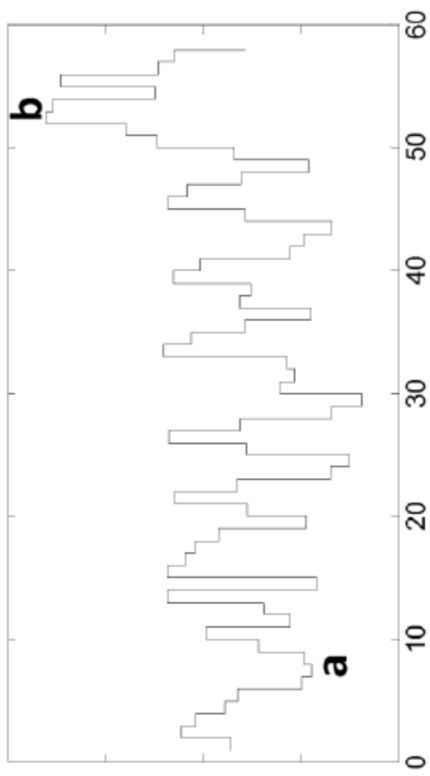


Figura 17A

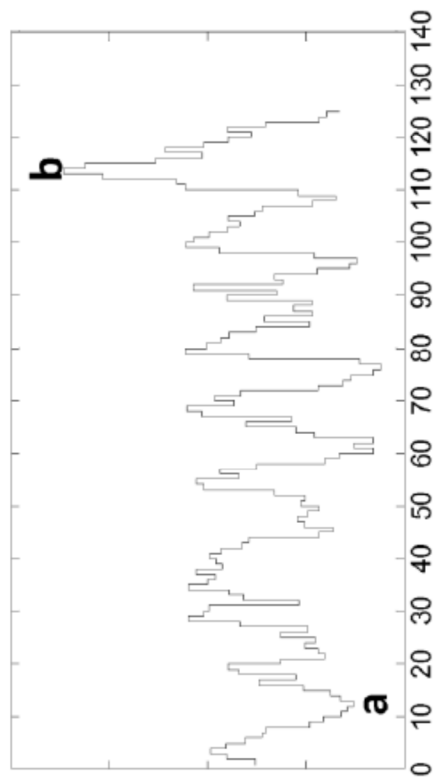


Figura 17C

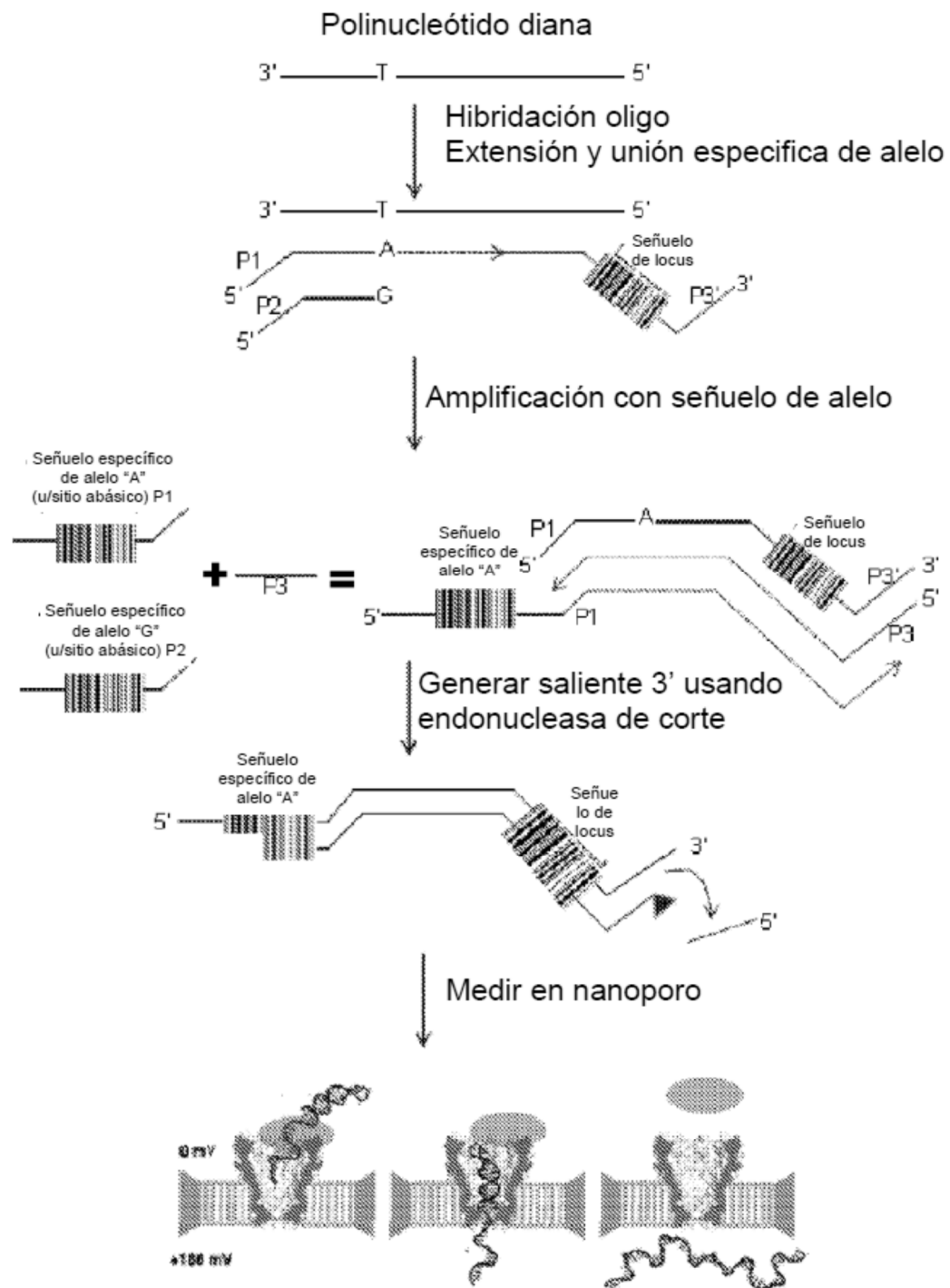


Figura 18

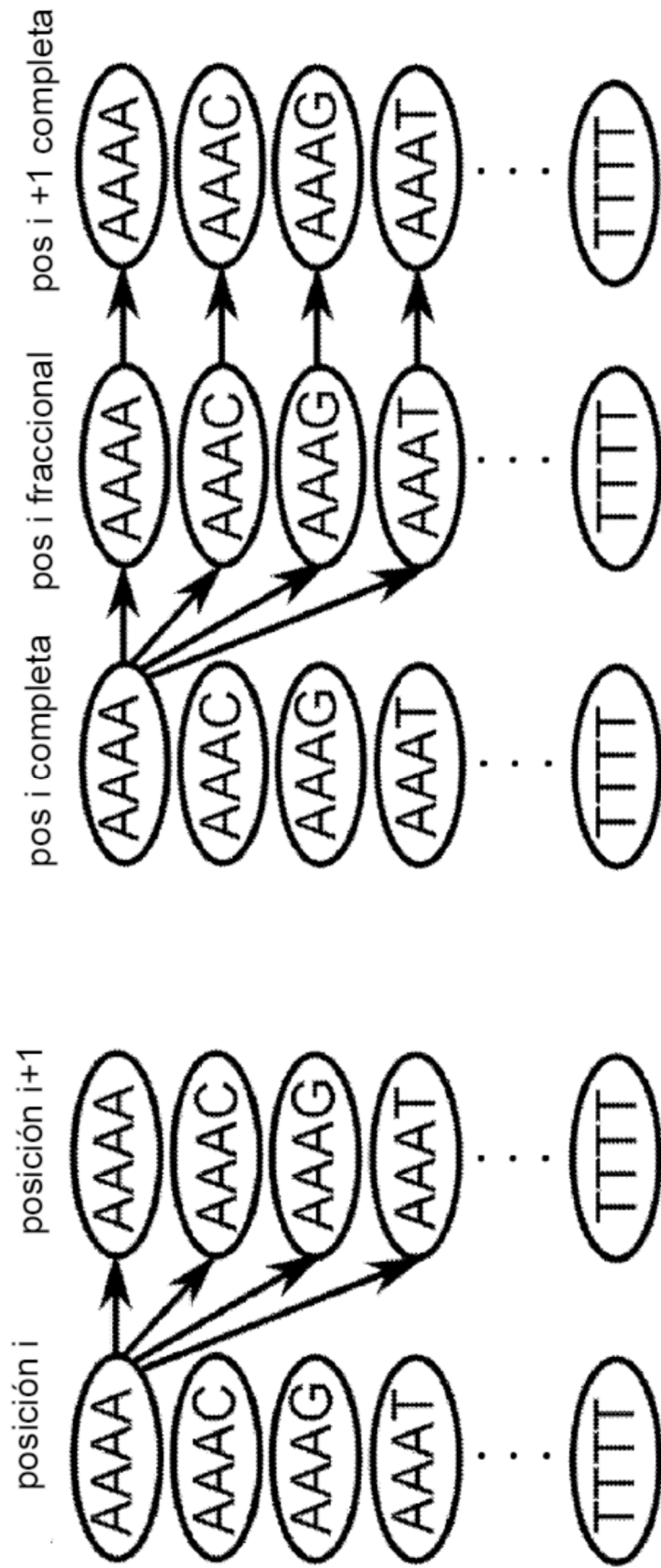


Figura 19A

Figura 19B

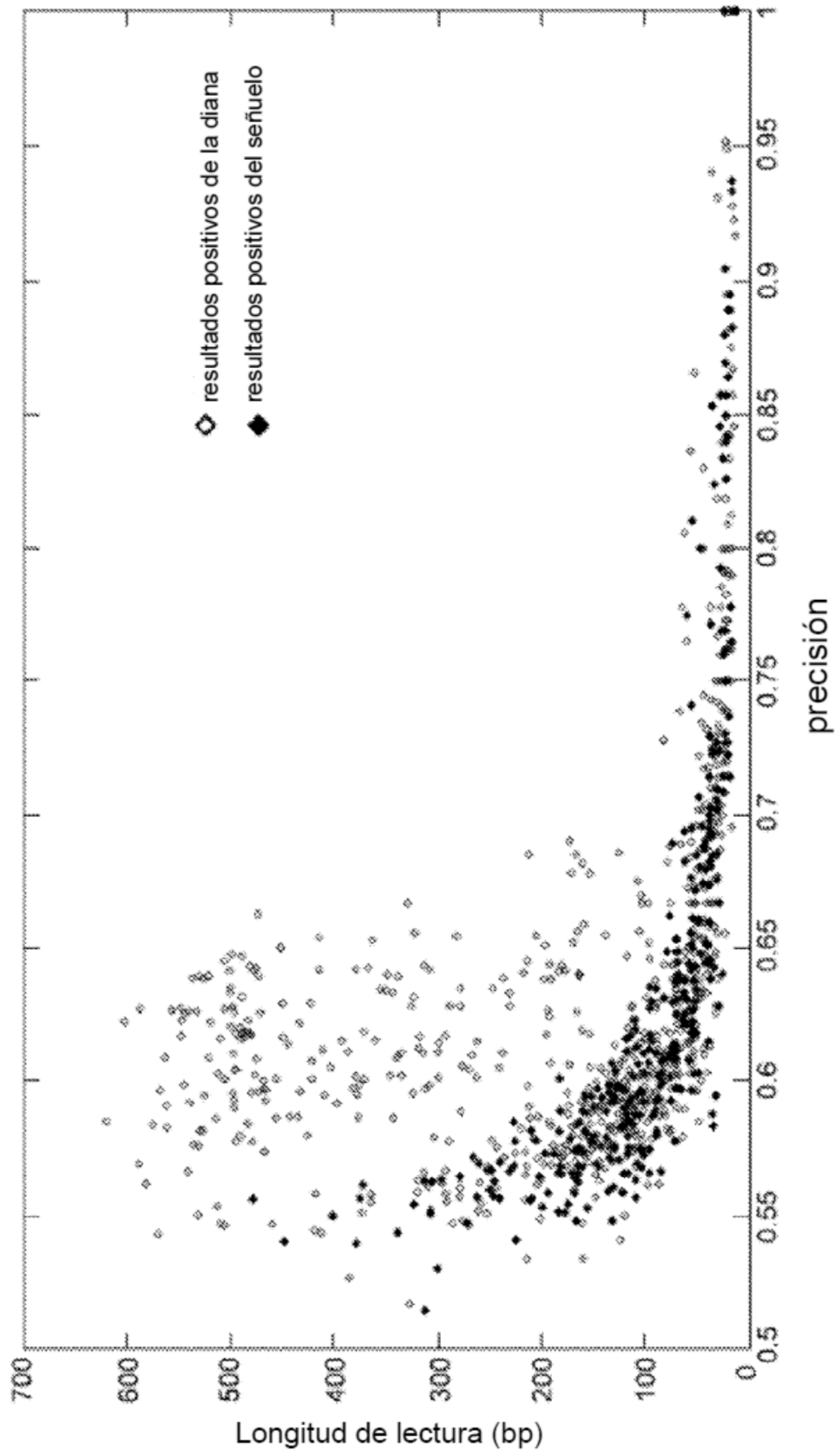


Figura 20A

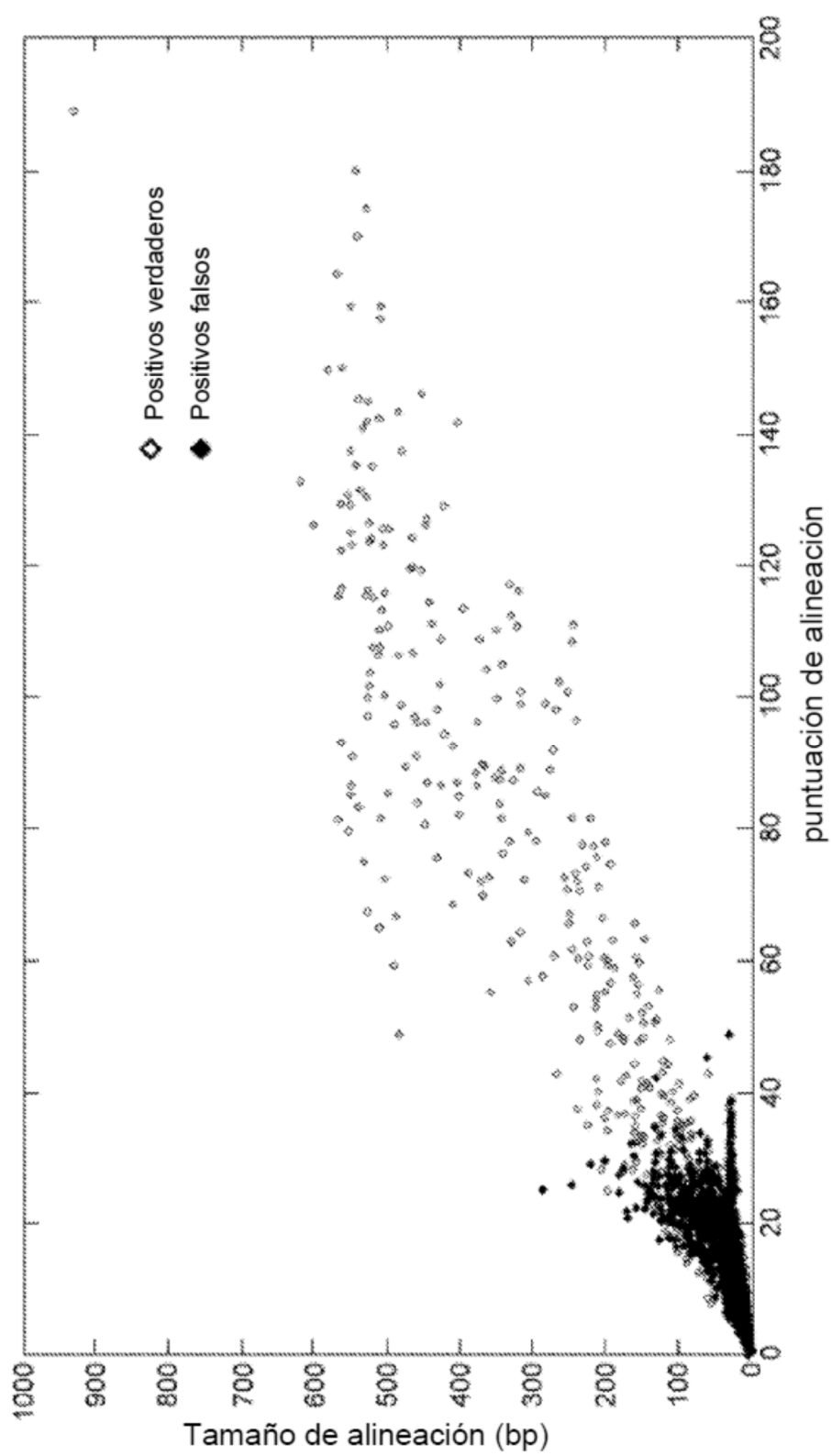


Figura 20B

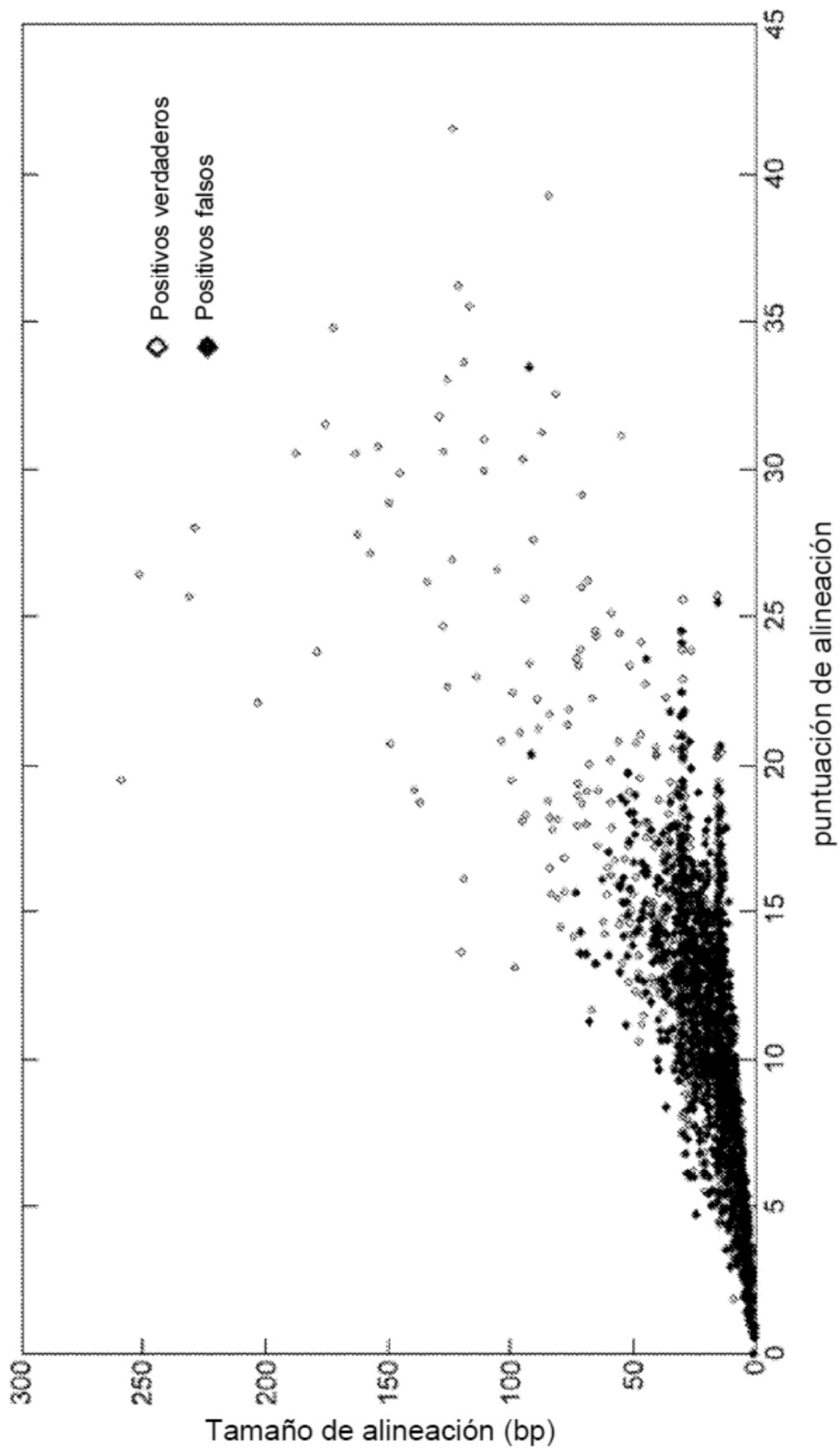


Figura 20C

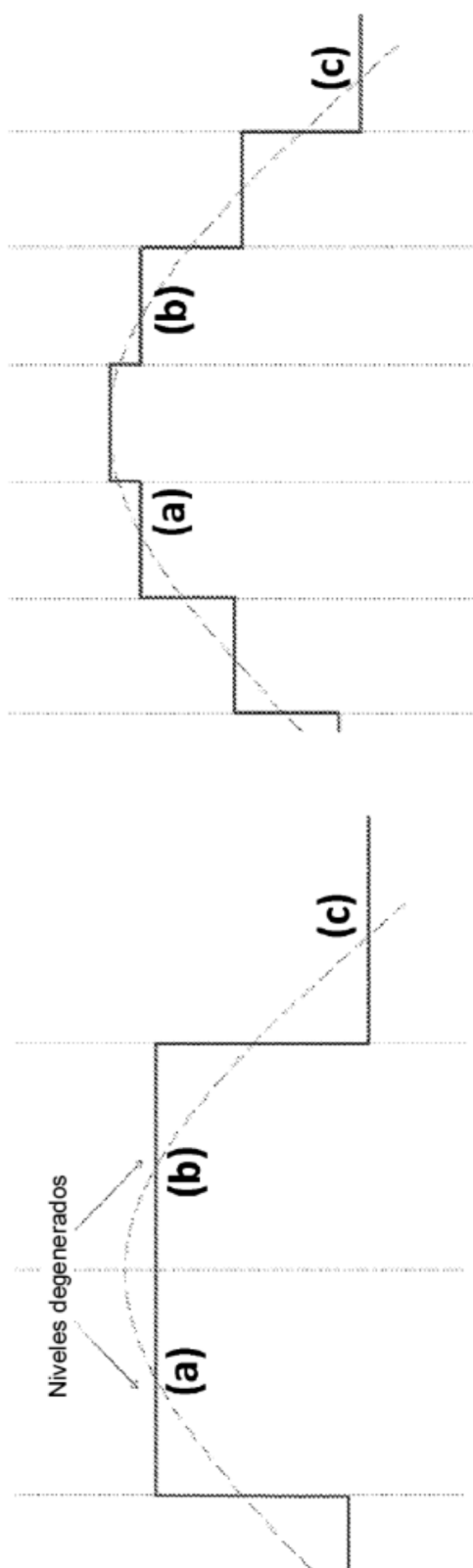


Figura 21A

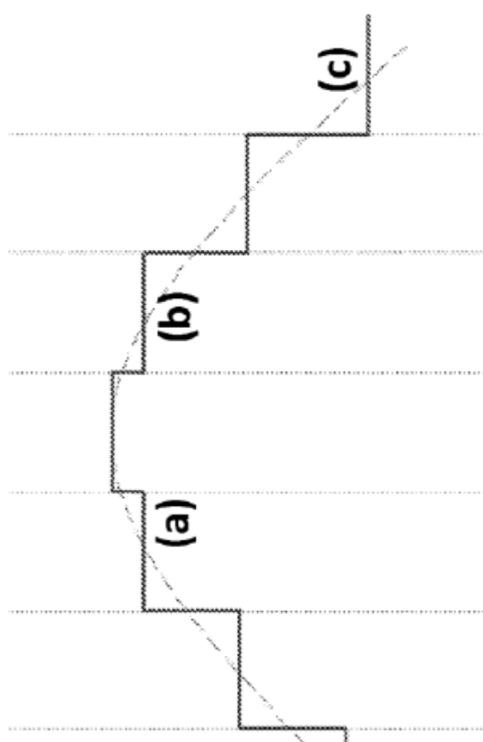


Figura 21B

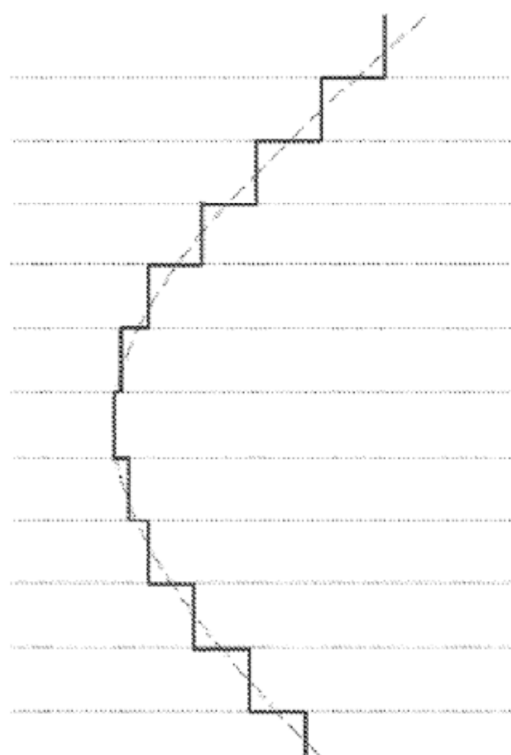


Figura 21C

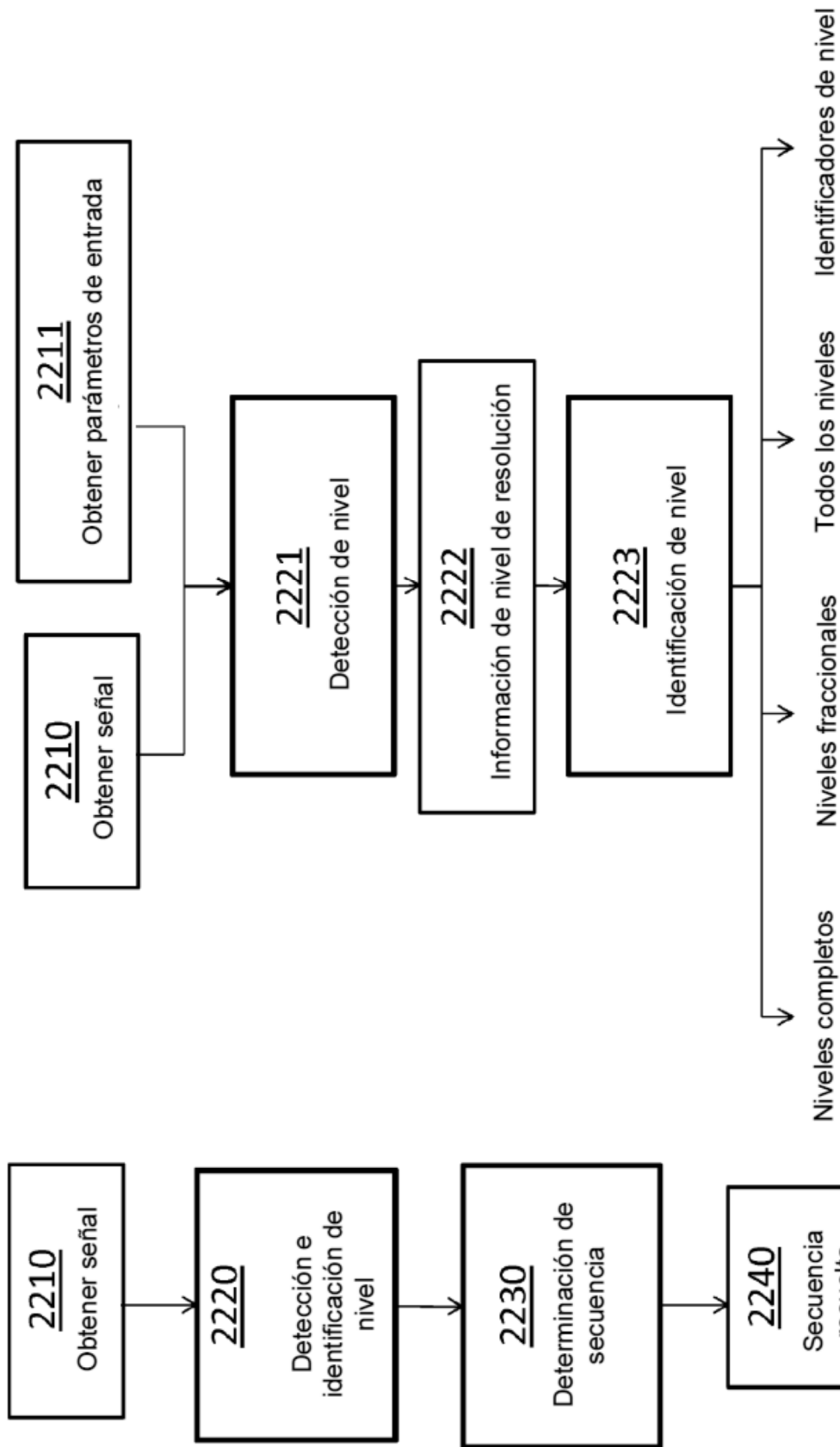


Figura 22A

Figura 22B

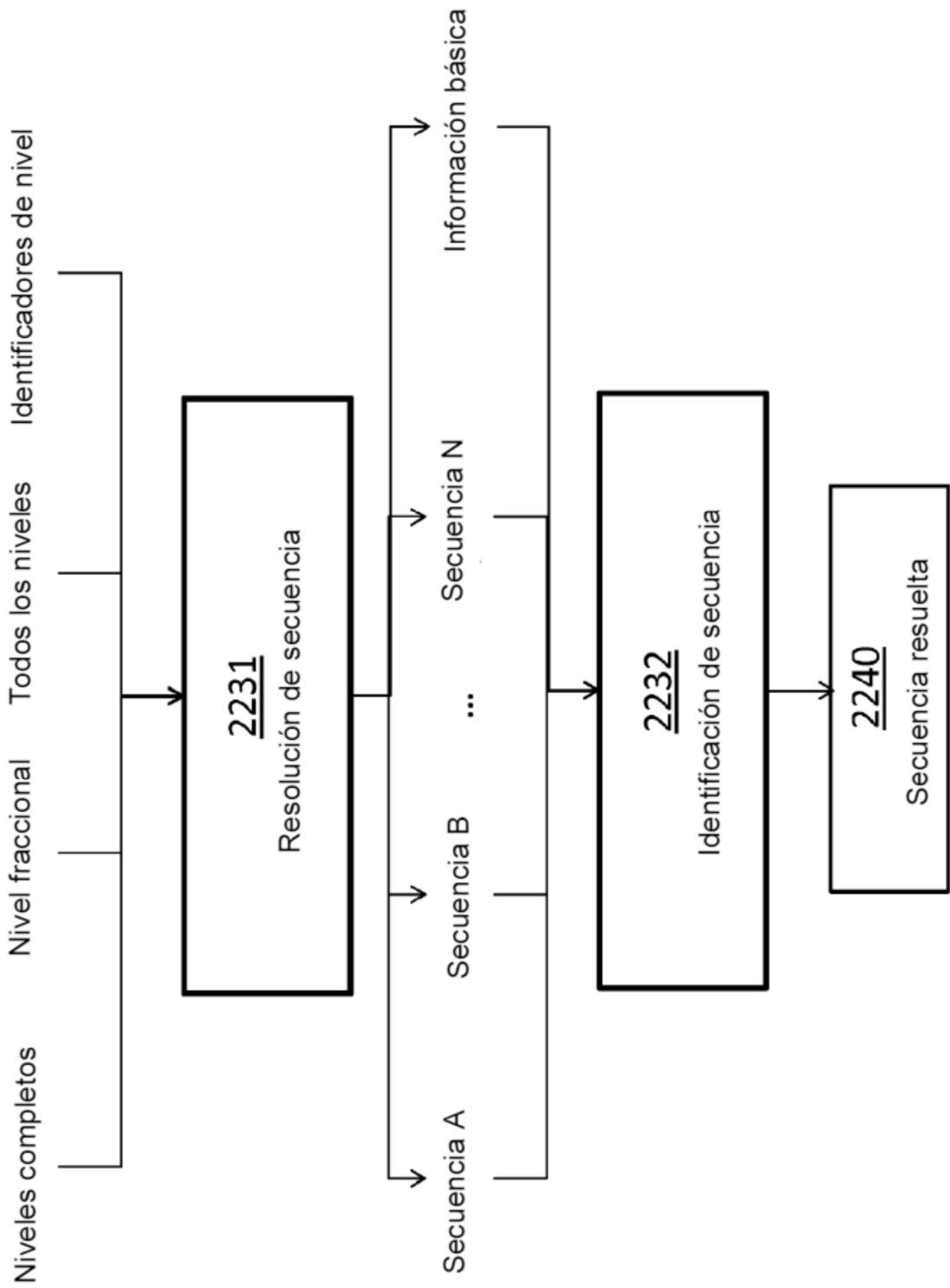


Figura 22C

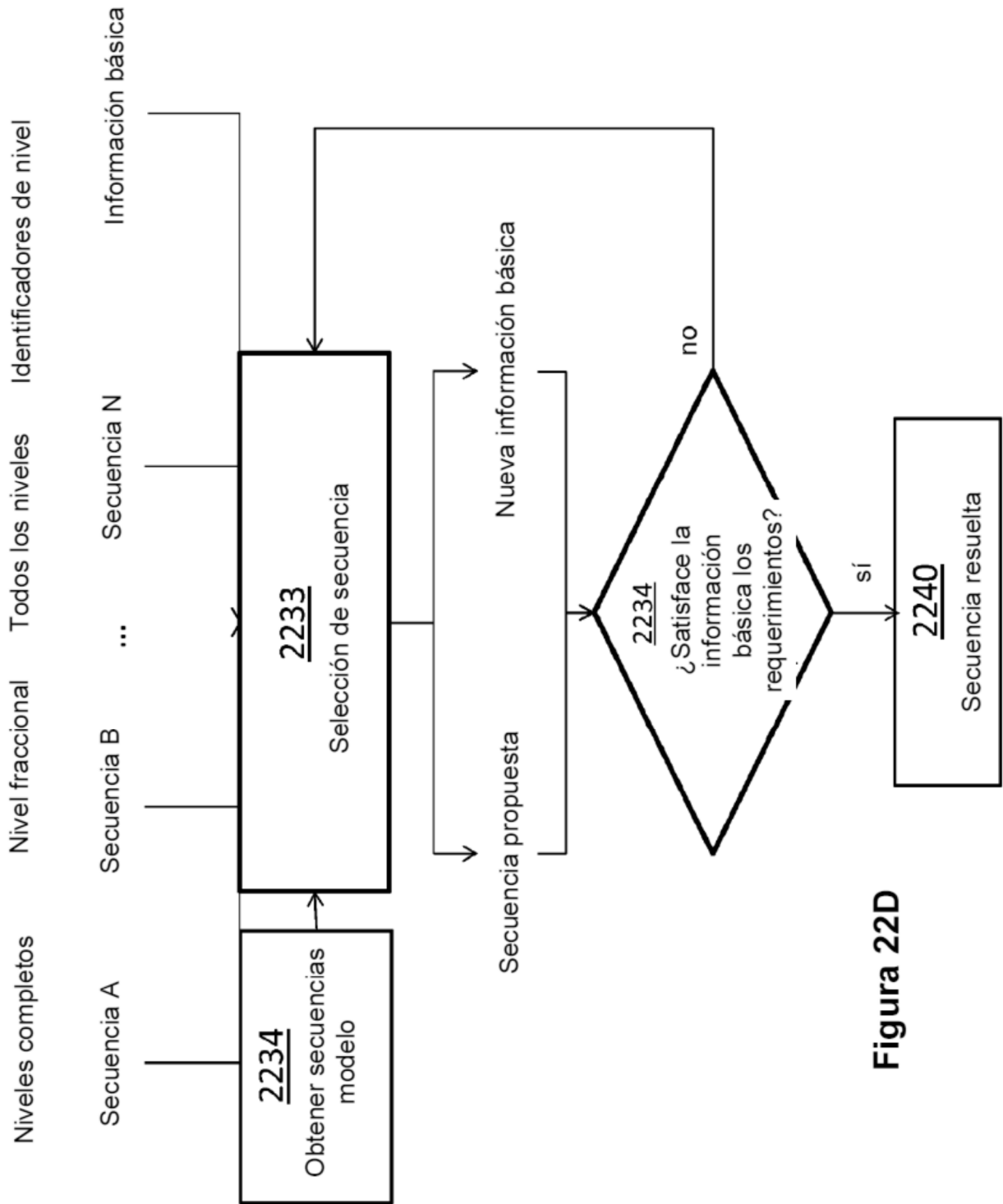


Figura 22D

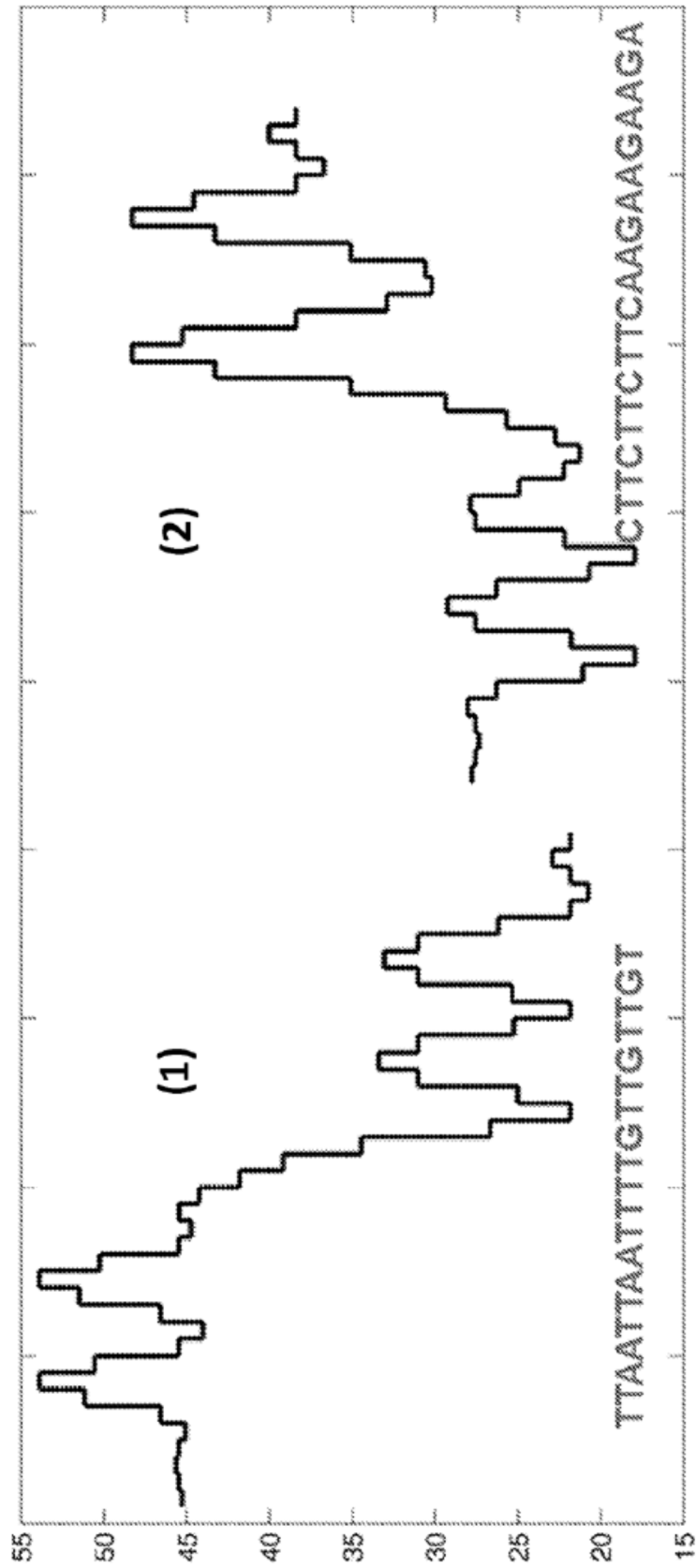


Figura 23

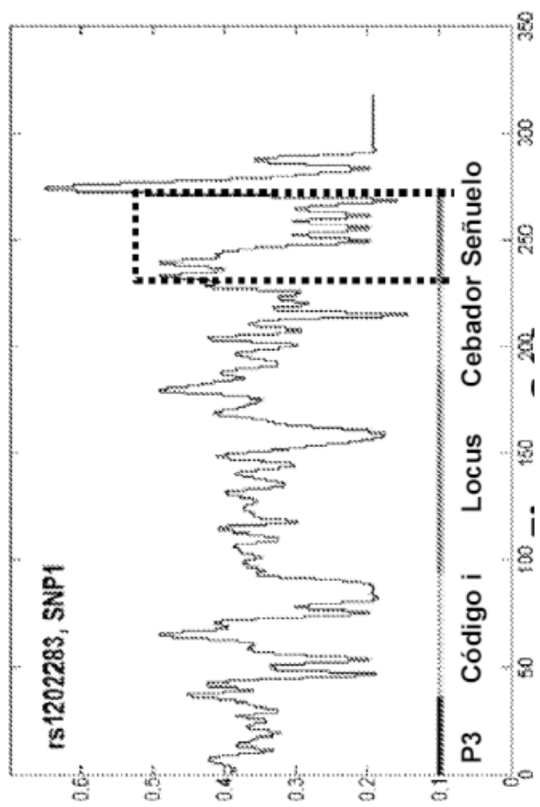


Figura 24A

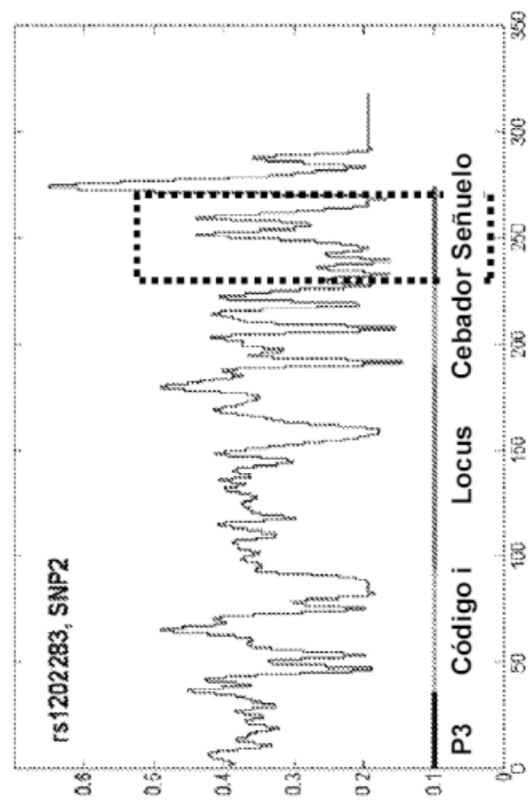


Figura 24C

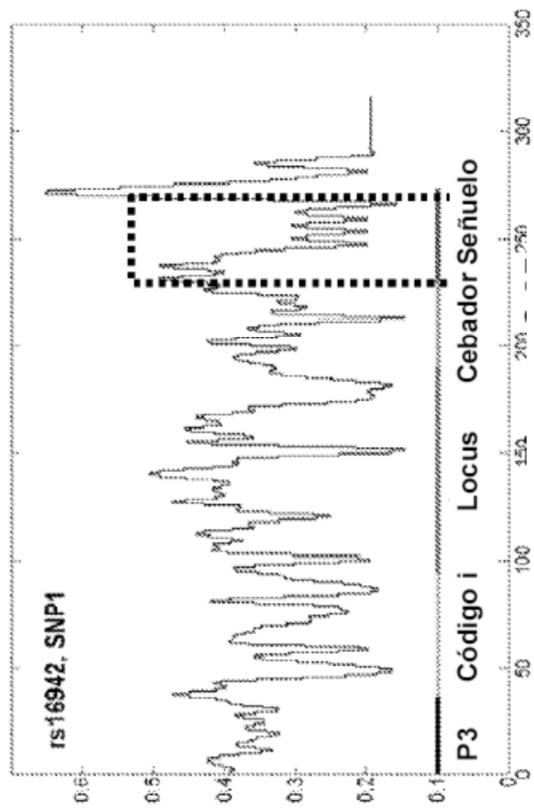


Figura 24B

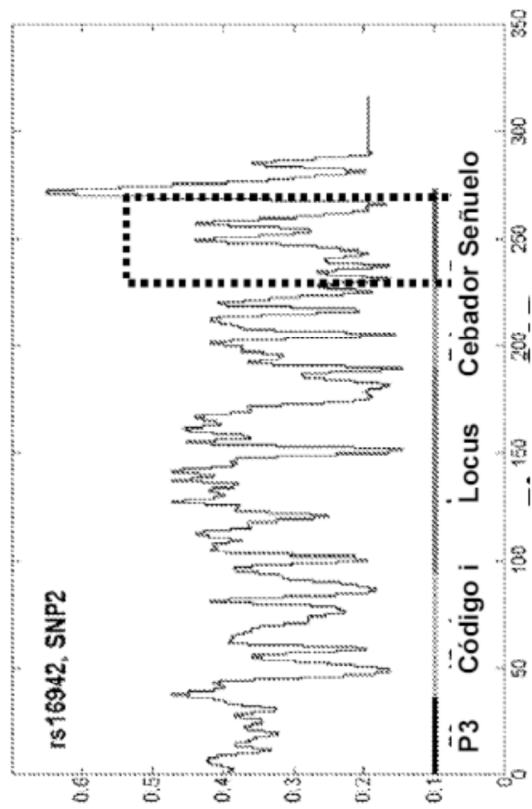


Figura 24D

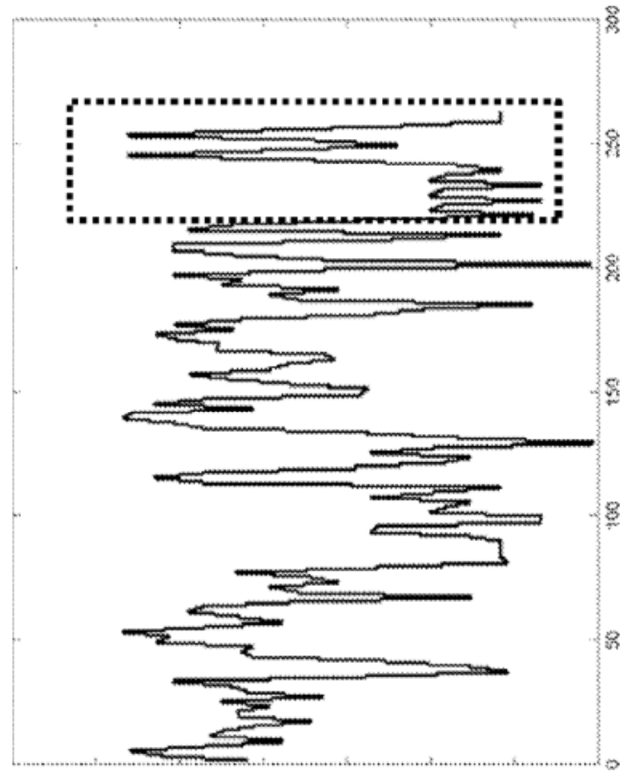


Figura 25B

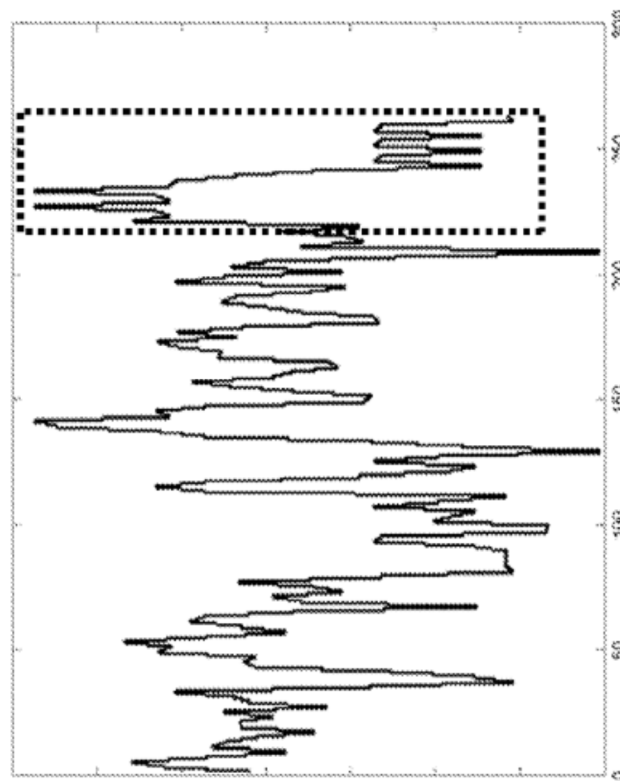


Figura 25A

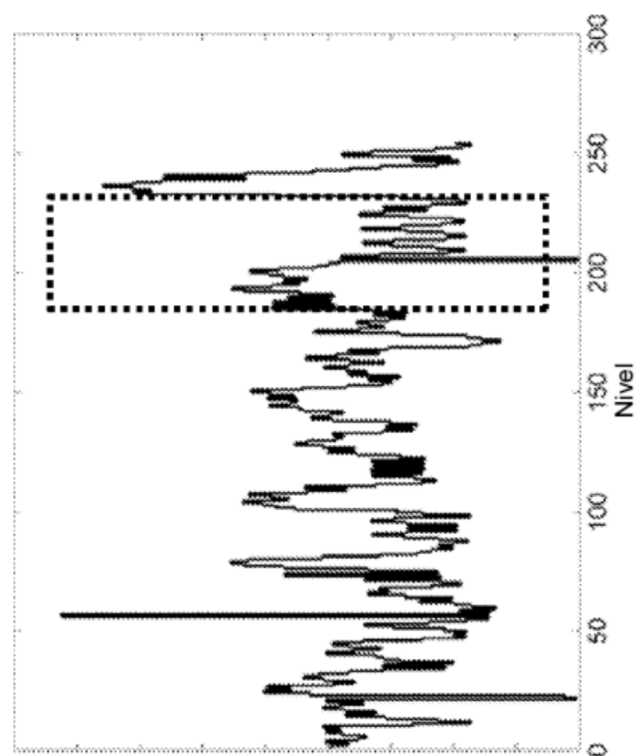


Figura 26B

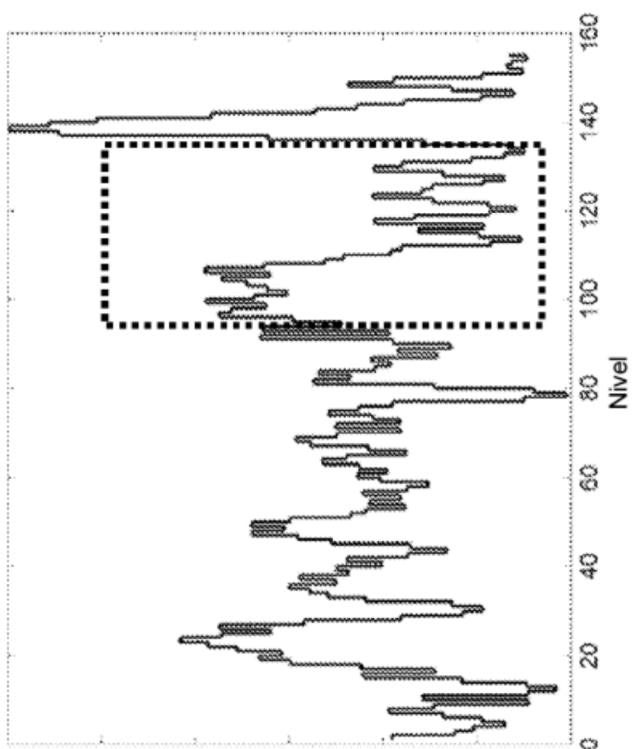


Figura 26A

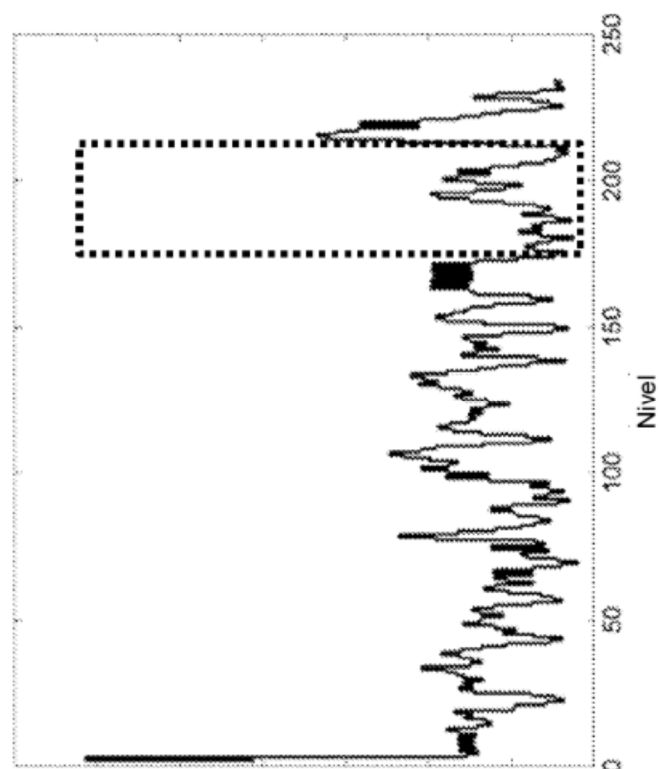


Figura 26D

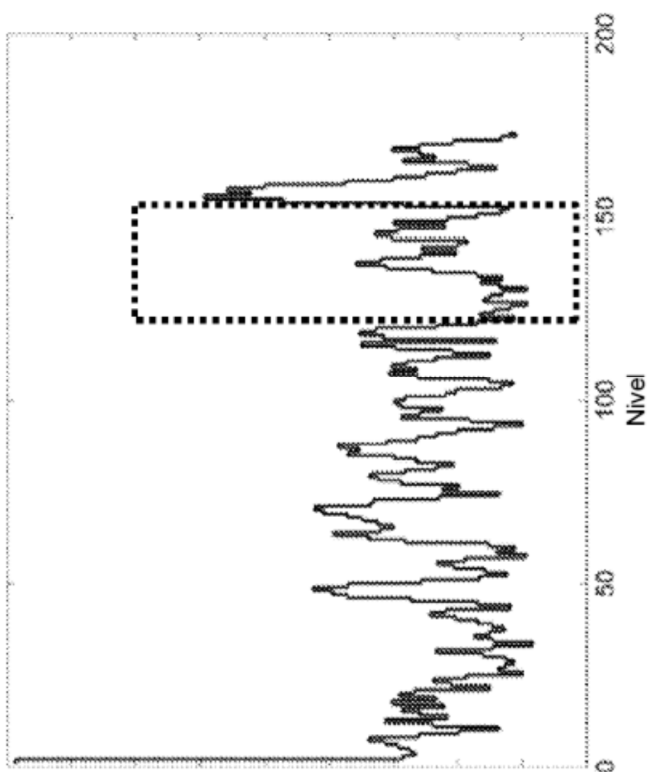


Figura 26C